

75776

ART UNIT 1642, ROOM 8A01, MB 8E12
305-2008

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

partly done 3/01

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: 116
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

DEFINITION RPCI-23-17L22-TV RPCI-23 Mus musculus genomic clone RPCI-23-17L22, DNA sequence.

ACCESSION AZ121142

VERSION AZ121142.1 GI:7787763

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C.M.

TITLE Mouse BAC End Sequences from Library RPCI-23

JOURNAL Unpublished (1999)

COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (inforesgen.com). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html

Plate: 17 row: L column: 22

Seq primer: '7'

Class: BAC ends.

Location/Qualifiers

1..183

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-17L22"

/clone_lib="RPCI-23"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI MethyIase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 65 a 29 c 24 g 65 t

ORIGIN

Query Match 7.5%; Score 20; DB 12; Length 183;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 tatgtttgtttgtttgtttgc 265
|||||

Db 27 TATGTTTGTGTTTGTTC 8

RESULT 15

BF290880/c

LOCUS

DEFINITION BF290880 184 bp mRNA linear EST 28-NOV-2000

EST455471 Rat Gene Index, normalized rat, Rattus norvegicus cDNA

Rattus norvegicus cDNA clone RGIIF59 3' sequence, mRNA sequence.

ACCESSION BF290880

VERSION BF290880.1 GI:11221950

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 184)

AUTHORS Malek, R.L., Cho, J., Lee, Y., Karamycheva, S., Parvizi, B., Pertea, G., Sultana, R., Tsai, J., White, J., Quackenbush, J., and Lee, N.H.

TITLE Generation of ESTs from Normalized Rat Embryo, Bento Soares Unpublished (2000)

JOURNAL Other ESTs: EST353021

COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org

This clone is available through the ATCC, contact the ATCC tel#703-365-2700 for further information.

Location/Qualifiers

1..184

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

/clone="RGIIF59"

/clone_lib="Rat Gene Index, normalized rat, Rattus norvegicus cDNA"

/tissue_type="mixed tissue"

/lab_host="DH5-alpha"

/note="Vector: pT377Pac; Site_1: EcoRI; Site_2: NotI; Combination of ROV, RBR, RKL, RLI, RPL, REM, RNU, RSP RHE, RPC, RPN"

BASE COUNT 80 a 34 c 26 g 44 t

ORIGIN

Query Match 7.5%; Score 20; DB 10; Length 184;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 gtatgtttgtttgtttgtttg 264
|||||

Db 71 GTATGTTTGTGTTTGTTC 52

Search completed: September 20, 2002, 08:34:02

Job time: 8726 sec

(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (inforesgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 384 row: 1 column: 12
 Seq primer: SP6
 Class: BAC ends.

FEATURES

Location/Qualifiers
 1..721
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-384I12"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI MethyIase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT

222 a 185 c 151 g 163 t

ORIGIN

Query Match 7.9%; Score 21; DB 12; Length 721;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 244 ggatgttttggttttgtttg 264
 |||||
 Db 224 GGTATGTTTGTGTTTGTGTTG 204

RESULT 12

AG014775

LOCUS

DEFINITION Homo sapiens genomic DNA, 723 bp DNA linear GSS 14-APR-1999 survey sequence.

ACCESSION

AG014775 AG006490

VERSION

AG014775.1 GI:3649993

KEYWORDS

GSS.

SOURCE

Homo sapiens DNA, clone: 762015N19.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 723)

AUTHORS

Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.

TITLE

Homo sapiens genomic DNA, chromosome 21q

JOURNAL

Published Only in DataBase (1998) In press

REFERENCE

2 (bases 1 to 723)

AUTHORS

Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.

TITLE

Direct Submission

JOURNAL

Submitted (23-SEP-1998) to the DDBJ/EMBL/GenBank databases.

Masahira Hattori, Kitasato University, Department of Science, JST

Sequencing Laboratory, Kitasato 1-15-1, Sagami-hara 228, Japan

(E-mail:hattori@hgc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,

Fax:0427-78-9501)

On Feb 6, 1999 this sequence version replaced gi:2992368.

AG006490: Submitted (27-Mar-1998).

COMMENT

Location/Qualifiers

1..723

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

/clone="762015N19"

/map="21q"

BASE COUNT

202 a 156 c 165 g 196 t 4 others

Query Match 7.9%; Score 21; DB 12; Length 723;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 ttcttcaagttaatttttcac 189

|||||

Db 514 TTTCTTCAAGTTAATTTTCAC 534

FEATURES

Location/Qualifiers

1..149

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-2A16"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 51 a 27 c 38 g 33 t

ORIGIN

Query Match 7.5%; Score 20; DB 12; Length 149;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 245 gtatgttttggttttgtttg 264
 |||||
 Db 33 GTATGTTTGTGTTTGTGTTG 14

FEATURES

Location/Qualifiers

1..149

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-2A16"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 51 a 27 c 38 g 33 t

ORIGIN

Query Match 7.9%; Score 21; DB 12; Length 723;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 169 ttcttcaagttaatttttcac 189
 |||||
 Db 514 TTTCTTCAAGTTAATTTTCAC 534

FEATURES

Location/Qualifiers

1..149

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-2A16"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 51 a 27 c 38 g 33 t

ORIGIN

Query Match 7.5%; Score 20; DB 12; Length 149;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 245 gtatgttttggttttgtttg 264
 |||||
 Db 33 GTATGTTTGTGTTTGTGTTG 14

FEATURES

Location/Qualifiers

1..149

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-2A16"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 51 a 27 c 38 g 33 t

ORIGIN

Query Match 7.9%; Score 21; DB 12; Length 723;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 169 ttcttcaagttaatttttcac 189
 |||||
 Db 514 TTTCTTCAAGTTAATTTTCAC 534

FEATURES

Location/Qualifiers

1..149

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-2A16"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 51 a 27 c 38 g 33 t

ORIGIN

Query Match 7.5%; Score 20; DB 12; Length 149;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 245 gtatgttttggttttgtttg 264
 |||||
 Db 33 GTATGTTTGTGTTTGTGTTG 14

FEATURES

Location/Qualifiers

1..149

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-2A16"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 51 a 27 c 38 g 33 t

ORIGIN

Query Match 7.9%; Score 21; DB 12; Length 723;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 169 ttcttcaagttaatttttcac 189
 |||||
 Db 514 TTTCTTCAAGTTAATTTTCAC 534

FEATURES

Location/Qualifiers

1..149

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-2A16"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 51 a 27 c 38 g 33 t

ORIGIN

Query Match 7.5%; Score 20; DB 12; Length 149;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 245 gtatgttttggttttgtttg 264
 |||||
 Db 33 GTATGTTTGTGTTTGTGTTG 14

FEATURES

Location/Qualifiers

1..149

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-2A16"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 51 a 27 c 38 g 33 t

ORIGIN

Query Match 7.9%; Score 21; DB 12; Length 723;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 169 ttcttcaagttaatttttcac 189
 |||||
 Db 514 TTTCTTCAAGTTAATTTTCAC 534

FEATURES

Location/Qualifiers

1..149

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-2A16"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 51 a 27 c 38 g 33 t

ORIGIN

Query Match 7.5%; Score 20; DB 12; Length 149;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 245 gtatgttttggttttgtttg 264
 |||||
 Db 33 GTATGTTTGTGTTTGTGTTG 14

FEATURES

Location/Qualifiers

1..149

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-2A16"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 51 a 27 c 38 g 33 t

ORIGIN

Query Match 7.9%; Score 21; DB 12; Length 723;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 169 ttcttcaagttaatttttcac 189
 |||||
 Db 514 TTTCTTCAAGTTAATTTTCAC 534

FEATURES

Location/Qualifiers

1..149

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-2A16"

/clone_lib="CHORI-230 Segment 1"

/sex="Female"

/cell_type="Brain"

/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 51 a 27 c 38 g 33 t

ORIGIN

Query Match 7.5%; Score 20; DB 12; Length 149;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 245 gtatgt

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (inforesgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 87 row: C column: 1
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers
1..557
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-87C1"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site: 1: EcoRI; Site: 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 179 a 133 c 109 g 136 t
ORIGIN

Query Match 7.9%; Score 21; DB 12; Length 557;

Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 244 ggtatgttttggttgtttgtttg 264
|||||
Db 208 GGTATGTTTGTGTTGTTGTTG 188

RESULT 10

AZ326781/c 614 bp DNA linear GSS 29-SEP-2000
LOCUS
DEFINITION
IM0049M16R Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGCLM0049M16 R, DNA sequence.

ACCESSION
AZ326781 GI:10384871
VERSION
GSS.

KEYWORDS
SOURCE
house mouse.

ORGANISM
Mus musculus

REFERENCE
AUTHORS
1 (bases 1 to 614)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL
COMMENT
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00

FEATURES

source

Plate: 0049 row: M column: 16
Seq primer: CACACAGAAACAGCTATGACC
Class: Plasmid ends
High quality sequence stop: 614.
Location/Qualifiers
1..614
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0049M16"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 209 a 132 c 89 g 184 t
ORIGIN

Query Match 7.9%; Score 21; DB 12; Length 614;

Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 244 ggtatgttttggttgtttgtttg 264
|||||
Db 487 GGTATGTTTGTGTTGTTGTTG 467

RESULT 11

AZ018710/c 721 bp DNA linear GSS 25-FEB-2000
LOCUS
DEFINITION
RPCI-23-384112.TJ RPCI-23 Mus musculus genomic clone RPCI-23-384112 , DNA sequence.

ACCESSION
AZ018710 GI:7094094
VERSION
GSS.

KEYWORDS
SOURCE
house mouse.

ORGANISM
Mus musculus

REFERENCE
AUTHORS
1 (bases 1 to 721)
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C. M.

TITLE
Mouse BAC End Sequences from Library RPCI-23

JOURNAL
COMMENT
Unpublished (1999)
Other_GSSs: RPCI-23-384112.TV
Contact: Shaying Zhao

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong

URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222,
Fax: 81-45-503-9216

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5', GAGAGAGAGCGCCCACTCGAGTTTCTTTTCTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence15',
GAGAGAGAGAGCTCAAGCTCAATTAATTAAACCCGCCCC 3']. cDNA was
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
XhoI. Host: SOLR.

FEATURES

source

Location/Qualifiers

1..4077
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGD:MG1:1904359"
/db_xref="taxon:10090"
/clone="1200016E19"
/sex="male"
/tissue_type="lung"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
14..2125
/gene="Nfatc1"
14..2125
/note="data source:MGD, source key:MG1:102469,
evidence:ISS
nuclear factor of activated T-cells, cytoplasmic 1
putative"

gene

CDS

/codon_start=1
/protein_id="BAB23581.1"
/db_xref="GI:12836271"

/translation="MTGLRDPDFDFLFEFDQSGGAAAEHXYVSPVSTLPL
PTAHSALPAACHDLOTSTPGISAVPSANHPYSGGAVDSGSCYFLSSNTRPNGAPT
LESPLREITSYGLHGGSCQFHDVEVDYLPCKRSPSTAILHPLSLRAYRDPSCLS
PASSLSRSCSEASSIESNYSPYASPTSPQSPCVSPKPTDPEEGRSLGACHL
LGSPPHSPTSPRASITEESWLGARGSRPTSPCKRKYSLNGRQPCSPHSPPTSPH
GSPRVSTEDTWLNTDYSALVAAINALATDSTLDLGDGVPKSKRTALEHAPSV
ALKVPAGDALDQLPSHSGPYELRIEVOQPKSHRAHYETEGSRGAVKASAGGPIVOL
SYMSPSLPDLQPSHSGPYELRIEVOQPKSHRAHYETEGSRGAVKASAGGPIVOL
HGYLENEPTLQFLFGTADRLRLPHAFYQVHRITGKTVSTTSHEILLNTRKVLLEPL
LPENMRAIIDCGILKLRNSDIELRKGEDTIGRKNTRVRLVFRVHIPQPNKRTLSLQ
VASNPTECSQSAQELPLVEKQSTDSYPVIGGKMYLGNHFLQDSKVTVEKAPDGH
HYWEMAKTDRLCKPNSLVVEIPFPNRQITSPVQVSPYVNCGRKRKRSQYORFTYLP
ANGNSVFLTSSSELRGCFY"

polyA_signal

4053..4058

/note="putative"

polyA_site

/note="putative"

BASE COUNT 944 a 1184 c 964 g 985 t

ORIGIN

Query Match 8.3%; Score 22; DB 11; Length 4077;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 aggtatgtttttgtttgtttg 264

|||||

Db 2189 AGGTATGTTTGTGTTTGTG 2168

RESULT 8

A0529197/c

LOCUS

DEFINITION

357 bp DNA linear GSS 18-MAY-1999
RPCI-11-367B5.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-367B5,
DNA sequence.

ACCESSION

A0529197

VERSION

A0529197.1

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 357)

AUTHORS

Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter

TITLE

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

JOURNAL

Map Building

COMMENT

Unpublished (1997)

Other GSSs: RPCI-11-367B5.TJ

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or fromResearch Genet cs (inforesgen.com). BAC end search page:http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..357

/organism="Homo sapiens"

/db_xref="GDB:7640572"

/db_xref="taxon:9606"

/clone="RPCI-11-367B5"

/clone_lib="RPCI-11"

/sex="Male"

/cell_type="Lymphocytes"

/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;

RPCI11 Human Male BAC Library"

BASE COUNT 113 a 97 c 77 g 70 t

ORIGIN

Query Match 7.9%; Score 21; DB 12; Length 357;

Best Local Similarity 100.0%; Pred. No. 3.9;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 aggtatgtttgtttgtttgtttt 263

|||||

Db 56 AGGTATGTTTGTGTTTGTGTTT 36

RESULT 9

A0529197/c

LOCUS

DEFINITION

557 bp DNA linear GSS 15-JUN-2000

RPCI-23-87C1.TJ RPCI-23 Mus musculus genomic clone RPCI-23-87C1,

DNA sequence.

ACCESSION AZ242621

VERSION AZ242621.1

KEYWORDS GI:8555812

SOURCE GSS.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 557)

AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret

,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.

and Fraser,C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other GSSs: RPCI-23-87C1.TJ

Contact: Shaying Zhao

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-A01-aej-d-05-0-UI"
/clone_lib="NIH_BMAP.MPG_N"
/dev_stage="27-32 days"
/db_host="DH10B (Life Technologies)"
/notes="Vector: pYT3D-Pac (Pharmacia) with a modified
polylinker; Site1: Not I; Site2: Eco RI; The
NIH_BMAP.MPG_N library is a normalized library
from mouse pineal gland. The tag is a string of 5
nucleotides present between the Not I site and the
oligo-dr track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
, 1996. Tissue provided by Ms. Annie Novakovich,
Zivic-Miller Laboratories.
TAG_LIB=NIH_BMAP.MPG_N
TAG_TISSUE=pineal-glands
TAG_SEQ=CAGAC"
BASE COUNT      84 a  68 c  96 g  137 t
ORIGIN

Query Match      8.3%; Score 22; DB 9; Length 385;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 242 taggtatgtttgtttgtttt 263
|||||
Db 331 TAGGTATGTTTGTGTTT 352

RESULT 6
AG014774      729 bp DNA linear GSS 14-APR-1999
LOCUS
DEFINITION
Homo sapiens genomic DNA, 21q region, clone: 762015N19, genomic
survey sequence.
ACCESSION
AG014774 AG006489
VERSION
AG014774.1 GI:3649992
KEYWORDS
GSS.
SOURCE
Homo sapiens DNA, clone:762015N19.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 729)
Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
Homo sapiens genomic DNA, chromosome 21q
Published Only in Database (1998) In press
REFERENCE
2 (bases 1 to 729)
Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
Direct Submission
Submitted (23-SEP-1998) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, Kitasato University, Department of Science, JST
Sequencing Laboratory; Kitasato 1-15-1, Sagamihara 228, Japan
(E-mail:hattori@qcc.ims.u-tokyo.ac.jp, Tel:0427-78-9732,
Fax:0427-78-9561)
On Feb 6, 1999 this sequence version replaced gi:2992367.
AG006489: Submitted (27-Mar-1998).
FEATURES
source
1..729
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="762015N19"
/map="21q"
BASE COUNT      208 a  153 c  153 g  203 t  12 others
ORIGIN

Query Match      8.3%; Score 22; DB 12; Length 729;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-A01-aej-d-05-0-UI"
/clone_lib="NIH_BMAP.MPG_N"
/dev_stage="27-32 days"
/db_host="DH10B (Life Technologies)"
/notes="Vector: pYT3D-Pac (Pharmacia) with a modified
polylinker; Site1: Not I; Site2: Eco RI; The
NIH_BMAP.MPG_N library is a normalized library
constructed
from mouse pineal gland. The tag is a string of 5
nucleotides present between the Not I site and the
oligo-dr track. The library was constructed as described
by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
, 1996. Tissue provided by Ms. Annie Novakovich,
Zivic-Miller Laboratories.
TAG_LIB=NIH_BMAP.MPG_N
TAG_TISSUE=pineal-glands
TAG_SEQ=CAGAC"
BASE COUNT      84 a  68 c  96 g  137 t
ORIGIN

Query Match      8.3%; Score 22; DB 9; Length 385;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 242 taggtatgtttgtttgtttt 263
|||||
Db 331 TAGGTATGTTTGTGTTT 352

RESULT 7
AK004810/c
LOCUS
DEFINITION
Mus musculus adult male lung cDNA, RIKEN full-length enriched
library, clone:1200016E19:nuclear factor of activated T-cells,
cytoplasmic 1, full insert sequence.
ACCESSION
AK004810
VERSION
AK004810.1 GI:12836270
KEYWORDS
Htc; CAP trapper.
SOURCE
Mus musculus (strain:C57BL/6J) adult male lung cDNA to mRNA,
clone:1200016E19.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (sites)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL
99279253
MEDLINE
PUBMED
10349636
AUTHORS
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL
20499374
MEDLINE
11042159
PUBMED
AUTHORS
Shibata,K., Aizawa,K., Nagao,K., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Ozawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multipillarary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL
20530913
MEDLINE
11076861
PUBMED
AUTHORS
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
REFERENCE
5 (bases 1 to 4077)
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hangaki,T.,
Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
Saito,H., Saito,K., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohana Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,

```

410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
 Tel: 780 492 0169
 Fax: 780 492 4265
 Email: smoores@afns.ualberta.ca
 Insert Length: 558 Std Error: 0.00
 POLYA-No.

FEATURES

source

Location/Qualifiers

1..558
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="Bos taurus Duodenum #1 library"
 /tissue_type="Smooth muscle"
 /cell_type="Simple columnar epithelial"
 /dev_stage="Young adult"
 /lab_host="XL1-BlueMRF/strain"
 /note="Organ: Intestine/Duodenum; Vector: Uni-2ZAPXR;
 Site_1: EcoRI; Site_2: Xho I"
 152 a 109 c 108 g 189 t

BASE COUNT

ORIGIN

Query Match 9.4%; Score 25; DB 10; Length 558;

Best Local Similarity 100.0%; Pred. No. 0.036;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 62 ctcatgccattatttttgacttc 86

Db 377 CTCATGCCATTATTTTATGACTTC 401
 |||||

RESULT

BM219592

LOCUS

DEFINITION BM219592 585 bp mRNA linear EST 31-JAN-2002
 Library (Long) Mus musculus cDNA clone C0927C09 3', mRNA sequence.

ACCESSION BM219592

VERSION BM219592.1 GI:17779452

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 585)

Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Luo, A.

and Ko, M.S.H.

Systematic Analyses of NIA Mouse 12.5-dpc Male Genital

Ridge/Mesonephros cDNA Library (Long)

Unpublished (2001)

CONTACT: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@igsun.grc.nia.nih.gov

Plate: C0927 row: C column: 09

Seq primer: -21M13 Forward

High quality sequence stop: 585

POLYA-Yes.

FEATURES

source

Location/Qualifiers

1..585
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="niaEST:C0927C09-3"
 /db_xref="taxon:10090"
 /clone="C0927C09"
 /clone_lib="NIA Mouse 12.5-dpc Male Genital
 Ridge/Mesonephros cDNA Library (Long)"
 /sex="Male"
 /tissue_type="Male genital ridge/mesonephros"
 /dev_stage="12.5-dpc"
 /lab_host="DH10B"
 /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
 NotI; Mouse cDNA project by the Laboratory of Genetics,
 National Institute on Aging (NIA), Intramural Research

Program, NIH (<http://igsun.grc.nia.nih.gov/cDNA>). This is
 a long-transcript enriched cDNA library (Ref. Genome Res.
 11: 1553-1558 (2001). [PMID: 11544199]). Double-stranded
 cDNAs were synthesized with an Oligo(dT) primer
 [Invitrogen]:

5'-pGACTAGTTCTAGATCGCGCGCCGCTTTT-3' from
 1.8 ug of total RNA, treated with T4 DNA polymerase, and
 purified by ethanol-precipitation. The cDNAs were ligated
 to Lone-linker LL-SalI, purified by phenol/chloroform, and
 separated from free linkers by Centricon 100. Then, the
 cDNAs were amplified by long-range high fidelity PCR using
 Ex Taq polymerase (Takara) with a primer SalI-S. The
 products were purified by phenol/chloroform and Centricon
 100. The cDNAs were digested with SalI and NotI enzymes
 and cloned into SalI/NotI site of pSPORT1 plasmid vector.
 The DH10B E. coli host was transformed with the ligation
 mixture by the standard chemical method. The average
 insert size is about 2.4 kb. The library was constructed
 by Yulan Piao (NIA)."

BASE COUNT 150 a 120 c 106 g 209 t

ORIGIN

Query Match 9.4%; Score 25; DB 10; Length 585;

Best Local Similarity 100.0%; Pred. No. 0.036;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 240 tttagtgatgtttttttttttttg 264

Db 73 TTTAGTATGTTTGTGTTTGTG 97
 |||||

RESULT

BM219592

LOCUS

DEFINITION BM219592 385 bp mRNA linear EST 15-JUL-1999
 UI-M-AOI-aej-d-05-0-UI-sl NIH_BMAP_MPG_N Mus musculus cDNA clone
 UI-M-AOI-aej-d-05-0-UI 3', mRNA sequence.

ACCESSION AI845434

VERSION AI845434.1 GI:5489335

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 385)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

CONTACT: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mEST@mail.nih.gov

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to verify it as a clone from the

normalized pineal glands library cDNA Library preparation: M.B.

Soares Lab Clone distribution: NIH BMAP cDNA clones will be made

available by the means that is soon to be determined. When NIH

determines the means for distribution of the BMAP cDNA clones, this

record will be updated accordingly when that means is determined.

The following repetitive elements were found in this cDNA sequence:

56-129, >B2#SINE/B2

Seq primer: M13 Forward

POLYA-res.

Location/Qualifiers

1..385

FEATURES

source

Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001).
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

e mouse tissues.

Location/Qualifiers

. 1. .500
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="T120429L21"
/clone_lib="RIKEN full-length enriched, 2 days neonate
sympathetic ganglion"
/sex="mixed"
/issue_type="sympathetic ganglion"
/dev_stage="#2 days neonate"
/lab_host="DH10B"
/note="Site.1: Sali; Site.2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAAGATCCAGAGCTCTTTTTTTTTTTN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by the
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGATTTCGAGTAATAAATAATCCCACCCCC 3']. cDNA
was cleaved with XhoI and BamHI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I.
Cloning sites, 5' end: Sali; 3' end: BamHI. Host: DH10B.
~RNA was provided by Akira Nakagawara, Div. Of
Biochemistry, Chiba Cancer Center Research Institute,
666-2 Nitona, Chuoh-ku, Chiba, 260-8717 Japan, whose
assistance we gratefully acknowledge."

BASE COUNT 181 A 95 C 94 G 130 T

ORIGIN

Query Match 9.4%; Score 25; DB 9; Length 500;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 tttaggtatggttttggtttggttg 264
 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dd 432 TTTAGGTATGTTTGTTGGTTTGTTC 408

RESULT 3

BM431438

LOCUS

DEFINITION

BW431438 Bos taurus Duodenum #1 library Bos taurus CDNA, mRNA linear EST 31-JAN-2002

ACCESSION

VERSION

KEYWORDS

SOURCE

COW.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 558)
. P.M.K. and Moore.S.S.,
Hansen,C., Fu,A., Meng,Y., Li,C., Okine,E., Sensen,C.W., Gordon
Gene Expression Profiling of the Bovine Gastrointestinal Tract
Unpublished (2002)

JOURNAL

TITLE

AUTHORS

REFERENCE

Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 08:33:58 ; Search time 5053.42 Seconds
(without alignments)
707.777 Million cell updates/sec

Title: US-09-802-520-9

Perfect score: 265

Sequence: 1 ccggagagggtcacagataa.....tatgtttgtttttgttc 265

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_htc.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_htc.*
- 12: gb_gss.*
- 13: em_gss_hum.*
- 14: em_gss_inv.*
- 15: em_gss_pln.*
- 16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	63	23.8	359	10 BF336378	BF336378 CM3-CT051
c 2	25	9.4	500	9 BB691606	BB691606 BB691606
c 3	25	9.4	558	10 BM431438	BM431438 lduo16f03
c 4	25	9.4	585	10 BM219592	BM219592 C0927C09-
c 5	22	8.3	385	9 A1845434	A1845434 UI-M-A01-
c 6	22	8.3	729	12 AG014774	AG014774 Homo sapi
c 7	22	8.3	4077	11 AK004810	AK004810 Mus muscu
c 8	21	7.9	357	12 AQ529197	AQ529197 RPCI-11-3
c 9	21	7.9	557	12 AZ242621	AZ242621 RPCI-23-8
c 10	21	7.9	614	12 AZ326781	AZ326781 IM0049M16
c 11	21	7.9	721	12 AZ018710	AZ018710 RPCI-23-3
c 12	21	7.9	723	12 AG014775	AG014775 Homo sapi
c 13	20	7.5	149	12 BH261930	BH261930 CH230-2A1
c 14	20	7.5	183	12 AZ121142	AZ121142 RPCI-23-1
c 15	20	7.5	184	10 BF290880	BF290880 EST455471
c 16	20	7.5	287	9 BB418291	BB418291 BB418291
c 17	20	7.5	315	12 AQ105175	AQ105175 HS_3000_A

18	20	7.5	325	9	BB550302	BB550302 BB550302
c 19	20	7.5	356	9	BE062468	BE062468 QV4-BT025
c 20	20	7.5	394	9	AI061309	AI061309 an32q03.x
c 21	20	7.5	397	9	A1429936	A1429936 vv36h12.y
c 22	20	7.5	398	12	AQ422494	AQ422494 RPCI-11-1
c 23	20	7.5	415	9	AW241285	AW241285 xa54d01.x
c 24	20	7.5	423	10	T85142	T85142 yd31g08.r1
c 25	20	7.5	429	12	AQ037959	AQ037959 CIT-HSP-2
c 26	20	7.5	433	9	AW836404	AW836404 PM3-LT003
c 27	20	7.5	445	12	AQ537589	AQ537589 RPCI-11-3
c 28	20	7.5	456	12	B94936	B94936 CIT-HSP-216
c 29	20	7.5	473	12	BH039744	BH039744 RPCI-24-3
c 30	20	7.5	483	12	BH039325	BH039325 CH230-129
c 31	20	7.5	503	12	AQ209876	AQ209876 HS_3242_B
c 32	20	7.5	504	10	W03237	W03237 za35b09.r1
c 33	20	7.5	520	10	BG791807	BG791807 UTSW_H1B1
c 34	20	7.5	541	9	AA755962	AA755962 vv36h12.r
c 35	20	7.5	566	10	BG791810	BG791810 UTSW_H1B2
c 36	20	7.5	568	12	AZ263527	AZ263527 RPCI-23-1
c 37	20	7.5	630	10	BG694791	BG694791 NISC-1v08
c 38	20	7.5	632	12	BH109251	BH109251 RPCI-24-3
c 39	20	7.5	632	12	AZ786548	AZ786548 2M0032B11
c 40	20	7.5	646	10	BG694722	BG694722 NISC-1v07
c 41	20	7.5	655	10	BG523183	BG523183 29-76 Ste
c 42	20	7.5	684	9	AV175132	AV175132 AV175132
c 43	20	7.5	690	12	AG184103	AG184103 Pan trogl
c 44	20	7.5	692	12	AG081865	AG081865 Pan trogl
c 45	20	7.5	789	12	AQ532149	AQ532149 RPCI-11-3

ALIGNMENTS

RESULT 1
BF336378/c
LOCUS BF336378 359 bp mRNA linear EST 22-NOV-2000
DEFINITION CM3-CT0510-010900-319-c09 CT0510 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF336378
VERSION BF336378.1 GI:11307126
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 359)
AUTHORS
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and
Simpson, A.J.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE
20202663
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, São Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-CT0510-
010900-319-c09&t3=2000-09-01&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 3
High quality sequence stop: 359.
Location/Qualifiers
1. 359
/organism="Homo sapiens"

FEATURES
source

```
Db 2316 TCTTTGTTTGTTC 2333
|||||
RESULT 11
US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

Query Match 6.8%; Score 18; DB 2; Length 246240;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 tgtttgtttgtttgttc 265
|||||
Db 88362 TCTTTGTTTGTTC 88379

RESULT 12
US-08-724-394A-21
; Sequence 21, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

Query Match 6.8%; Score 18; DB 2; Length 246240;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 tgtttgtttgtttgttc 265
|||||
Db 88362 TCTTTGTTTGTTC 88379

RESULT 13
US-08-724-394A-22
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-21

Query Match 6.8%; Score 18; DB 2; Length 246240;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 tgtttgtttgtttgttc 265
|||||
Db 88362 TCTTTGTTTGTTC 88379
```

; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; ORIGINAL SOURCE: human
 ; ORGANISM: human
 ; IMMEDIATE SOURCE: F55
 ; CLONE: F55
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 358 to 486
 ; LOCATION: 560 to 799
 ; LOCATION: 1042 to 1182
 ; LOCATION: 2105 to 2269
 ; LOCATION: 2370 to 2462
 ; IDENTIFICATION METHOD: by experiment
 ;
 US-08-909-965C-1

Query Match 6.8%; Score 18; DB 2; Length 2688;
 Best Local Similarity 100.0%; Pred. No. 4.2;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 248 tgtttgtttgtttgttc 265
 |||||
 Db 951 tgtttgtttgtttgttc 968

RESULT 9
 US-08-162-146-2
 ; Sequence 2, Application US/08162146
 ; Patent No. 5965788
 ; GENERAL INFORMATION:
 ; APPLICANT: HOUEBINE, Louis-Marie
 ; APPLICANT: DEVINOY, Eve
 ; TITLE OF INVENTION: Production of a Protein of Interest in
 ; TITLE OF INVENTION: the Milk of a Transgenic Mammalian
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; FILING DATE: 10-FEB-1994
 ; APPLICATION NUMBER: US/08/162,146
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/FR92/00533
 ; FILING DATE: 12-JUN-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: FR 91/07179
 ; FILING DATE: 12-JUN-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: WEGNER, Harold C.
 ; REGISTRATION NUMBER: 25,258
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 672-5300
 ; TELEFAX: (202) 672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4157 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; FEATURE:

; NAME/KEY: CDS
 ; LOCATION: Join(1868...1949, 2462...2587, 2888...3046, 3416
 ; LOCATION: ..3429)
 ;
 US-08-162-146-2

Query Match 6.8%; Score 18; DB 2; Length 4157;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 248 tgtttgtttgtttgttc 265
 |||||
 Db 2316 tgtttgtttgtttgttc 2333

RESULT 10
 US-09-314-127-2
 ; Sequence 2, Application US/09314127
 ; Patent No. 6268545
 ; GENERAL INFORMATION:
 ; APPLICANT: HOUEBINE, Louis-Marie
 ; APPLICANT: DEVINOY, Eve
 ; TITLE OF INVENTION: Production of a Protein of Interest in
 ; TITLE OF INVENTION: the Milk of a Transgenic Mammalian
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/314,127
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/162,146
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: FR 91/07179
 ; FILING DATE: 12-JUN-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: WEGNER, Harold C.
 ; REGISTRATION NUMBER: 25,258
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 672-5300
 ; TELEFAX: (202) 672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4157 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: Join(1868...1949, 2462...2587, 2888...3046, 3416
 ; LOCATION: ..3429)
 ;
 US-09-314-127-2

Query Match 6.8%; Score 18; DB 4; Length 4157;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 248 tgtttgtttgtttgttc 265

APPLICANT: OHL, Stephan Andreas
APPLICANT: SJMONS, Peter Christian
APPLICANT: VAN DER LEE, Frederique Marianne
APPLICANT: GODDUN, Oscar Johannes Maria
APPLICANT: Klap, Joke Johanna Catharina
TITLE OF INVENTION: NEMATODE-INDUCIBLE PLANT GENE PROMOTER
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: LADAS & PARRY
STREET: 26 WEST 61 STREET
CITY: NEW YORK
STATE: NY
ZIP: 10023-7604
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.25" Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS 95
SOFTWARE: WORDPERECT 8
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/117,927
FILING DATE: 26-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP96/02437
FILING DATE: 04-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: MASS, CLIFFORD J.
REGISTRATION NUMBER: 30,086
REFERENCE/DOCKET NUMBER: U-012011-9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 708-1890
TELEFAX: (212) 246-8959
TELEX: 233288
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2163 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Arabidopsis thaliana
STRAIN: C24
FEATURE:
NAME/KEY: CDS
LOCATION: 2161..2163
OTHER INFORMATION: /codon_start= 2161
FEATURE:
NAME/KEY: misc.feature
LOCATION: 2128..2163
OTHER INFORMATION: /note= "Sequence of pMOG553
Patent No. 6262344
OTHER INFORMATION: upstream (5') of the uid A translation initiation
OTHER INFORMATION: codon up to the RB/plant genome transition."
FEATURE:
NAME/KEY: promoter
LOCATION: 1..2127
FEATURE:
NAME/KEY: primer_bind
LOCATION: 787..804
OTHER INFORMATION: /label= primer6044-1
OTHER INFORMATION: /note= "annealing of primer 6044-1 (table 4) to
amplify subfragment"
FEATURE:
NAME/KEY: primer_bind
LOCATION: 1147..1169
OTHER INFORMATION: /label= primer6044-2
OTHER INFORMATION: /note= "annealing of primer 6044-2 (table 4) to
amplify subfragment"
FEATURE:
NAME/KEY: primer_bind
LOCATION: 1853..1880

OTHER INFORMATION: /label= primer6044-3
OTHER INFORMATION: /note= "annealing of primer 6044-3 (table 4) to
amplify subfragments"
FEATURE:
NAME/KEY: primer_bind
LOCATION: 1918..1940
OTHER INFORMATION: /label= primer6044-4
OTHER INFORMATION: /note= "annealing of primer 6044-4 (table 4) to
amplify subfragments"
FEATURE:
NAME/KEY: primer_bind
LOCATION: 1897..1917
OTHER INFORMATION: /label= primer6044-5
OTHER INFORMATION: /note= "annealing of primer 6044-5 (table 4) to
amplify subfragments (opposite strand)"
US-09-117-927-4
Query Match 6.8%; Score 18; DB 4; Length 2163;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 248 tggtttgtttgtttgtc 265
|||||
DB 853 TGTTCGTTTGTTCGTC 870
|||||
RESULT 8
US-08-909-965C-1
; Sequence 1, Application US/08909965C
; Patent No. 5936078
; GENERAL INFORMATION:
; APPLICANT: Kuga Tetsuo
; APPLICANT: Nakagawa Satoshi
; APPLICANT: Sakaki Yoshiyuki
; APPLICANT: Zhao Nanding
; APPLICANT: Hashida Hideji
; TITLE OF INVENTION: NOVEL DNA, NOVEL POLYPEPTIDE
; TITLE OF INVENTION: AND NOVEL ANTIBODY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FITZPATRICK, CELLA, HARPER AND SCINTO
; STREET: 277 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10172-0194
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/909,965C
; FILING DATE: August 12, 1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 322745/95
; APPLICATION NUMBER: PCT/JP96/03630
; FILING DATE: 12-No. 5936078-1995
; FILING DATE: 12-Dec-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Lawrence S. Perry
; REGISTRATION NUMBER: 31865
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-2400
; TELEFAX: 212-758-2982
; TELEX: 236262
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2688 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

NAME/KEY: unsure
LOCATION: 1424
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 1458
OTHER INFORMATION: unknown
NAME/KEY: CDS
LOCATION: (299)...(1336)
US-09-593-711A-3

Query Match 6.8%; Score 18; DB 4; Length 1910;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 247 atgtttgtttgtttg 264
|||||

Db 1770 atgtttgtttgtttg 1787

RESULT 5

US-07-601-094-1
Sequence 1, Application US/07601094
Patent No. 5215892

GENERAL INFORMATION:

APPLICANT: Kishimoto, Tadimitsu
APPLICANT: Hirano, Toshio
APPLICANT: Akira, Shizuo
APPLICANT: Isshiki, Hiroshi
APPLICANT: Tanabe, Osamu
APPLICANT: Kinoshita, Shigeml
APPLICANT: Shimamoto, Takuya
TITLE OF INVENTION: C/EBP2 Gene and Recombinant
TITLE OF INVENTION: C/EBP2
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESSES:

ADDRESSEE: Sughrue, Mion, Zinn, Macpeak &
ADDRESS: Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.

COUNTRY: United States

ZIP: 20037-3202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/601,094
FILING DATE: 19901022

CLASSIFICATION: 435

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1914 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 281..1316

OTHER INFORMATION:

US-07-601-094-1

Query Match 6.8%; Score 18; DB 1; Length 1914;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 247 atgtttgtttgtttg 264
|||||

Db 1752 ATGTTTGTGTTGTTG 1769

RESULT 6

US-08-012-735-1
Sequence 1, Application US/08012735
Patent No. 5360894

GENERAL INFORMATION:

APPLICANT: Kishimoto, Tadimitsu
APPLICANT: Hirano, Toshio
APPLICANT: Akira, Shizuo
APPLICANT: Isshiki, Hiroshi
APPLICANT: Tanabe, Osamu
APPLICANT: Kinoshita, Shigeml
APPLICANT: Shimamoto, Takuya
TITLE OF INVENTION: C/EBP2 Gene and Recombinant
TITLE OF INVENTION: C/EBP2
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESSES:

ADDRESSEE: Sughrue, Mion, Zinn, Macpeak &
ADDRESS: Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.

COUNTRY: United States

ZIP: 20037-3202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/012,735
FILING DATE: 19930203

CLASSIFICATION: 435

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1914 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 281..1316

OTHER INFORMATION:

US-08-012-735-1

Query Match 6.8%; Score 18; DB 1; Length 1914;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 247 atgtttgtttgtttg 264
|||||

Db 1752 ATGTTTGTGTTGTTG 1769

RESULT 7

US-09-117-927-4
Sequence 4, Application US/09117927
Patent No. 6262344

GENERAL INFORMATION:

RESULT 2

US-08-594-031-84/c
; Sequence 84, Application US/08594031
; Patent No. 5783182
; GENERAL INFORMATION:
; APPLICANT: THOMSON, Timothy C.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING METASTATIC SEQUENCES
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BAKER & BOTTS, L.L.P.
; STREET: 1299 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC

; COUNTRY: USA
; ZIP: 20004-2400
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/594,031
; FILING DATE: 30-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/006,838
; FILING DATE: 16-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Remenick, James
; REGISTRATION NUMBER: 36,902
; REFERENCE/DOCKET NUMBER: 0A146-0110
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-639-7700
; TELEFAX: 202-639-7890
; TELEX:

; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 252 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-594-031-84

Query Match 6.8%; Score 18; DB 1; Length 252;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 248 tgattgtttgtttgttc 265
|||||
Db 84 TGTGTTGTTGTTGTC 67

RESULT 3

US-09-018-584A-4
; Sequence 4, Application US/09018584
; Patent No. 6238863
; GENERAL INFORMATION:
; APPLICANT: Schumm, James W.
; APPLICANT: Bachter, Jeffery W.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR
; IDENTIFYING AND ANALYZING INTERMEDIATE TANDEM
; REPEAT DNA MARKERS
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Promega Corporation
; STREET: 2800 Woods Hollow Road
; CITY: Madison

; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53711-5399
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Mb
; COMPUTER: IBM compatible PC
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Word 97 (DOS text format)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,584A
; FILING DATE: 04-Feb-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Grady J. Frenchick
; REGISTRATION NUMBER: 29,018
; REFERENCE/DOCKET NUMBER: 16026.9180
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 257-3501
; TELEFAX: (608) 257-2275
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 317 bp
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Circular
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: no
; IMMEDIATE SOURCE:
; LIBRARY: plasmid, pGem3zf(+)
; CLONE: C331
US-09-018-584A-4

Query Match 6.8%; Score 18; DB 4; Length 317;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 247 atgtttgtttgtttgtt 264
|||||
Db 216 ATGTTTGTGTTGTTG 233

RESULT 4

US-09-593-711A-3
; Sequence 3, Application US/09593711A
; Patent No. 6271030
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Madeline M. Butler
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF C/EBP BETA EXPRESSION
; FILE REFERENCE: RTS-0118
; CURRENT APPLICATION NUMBER: US/09/593,711A
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 244
; SEQ ID NO 3
; LENGTH: 1910
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION:
; NAME/KEY: unsure
; LOCATION: 1415
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 1421
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 1422
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 1423
; OTHER INFORMATION: unknown

Oy 246 tatgttttggttttg 264
| | | | | | | | | | | | | | |
Db 194 tatgttttggttttg 212

Search completed: September 20, 2002, 10:00:07
Job time: 11725 sec

```
XX WO9839448-A2.
XX 11-SEP-1998.
XX 06-MAR-1998; 98WO-US04493.
XX 02-OCT-1997; 97US-0061060.
XX 07-MAR-1997; 97US-0038621.
XX 07-MAR-1997; 97US-0040161.
XX 07-MAR-1997; 97US-0040162.
XX 07-MAR-1997; 97US-0040163.
XX 07-MAR-1997; 97US-0040333.
XX 07-MAR-1997; 97US-0040334.
XX 07-MAR-1997; 97US-0040336.
XX 07-MAR-1997; 97US-0040626.
XX 11-APR-1997; 97US-0043311.
XX 11-APR-1997; 97US-0043312.
XX 11-APR-1997; 97US-0043313.
XX 11-APR-1997; 97US-0043314.
XX 11-APR-1997; 97US-0043568.
XX 11-APR-1997; 97US-0043569.
XX 11-APR-1997; 97US-0043570.
XX 11-APR-1997; 97US-0043578.
XX 11-APR-1997; 97US-0043580.
XX 11-APR-1997; 97US-0043669.
XX 11-APR-1997; 97US-0043670.
XX 11-APR-1997; 97US-0043671.
XX 11-APR-1997; 97US-0043672.
XX 11-APR-1997; 97US-0043674.
XX 23-MAY-1997; 97US-0047492.
XX 23-MAY-1997; 97US-0047501.
XX 23-MAY-1997; 97US-0047502.
XX 23-MAY-1997; 97US-0047503.
XX 23-MAY-1997; 97US-0047581.
XX 23-MAY-1997; 97US-0047582.
XX 23-MAY-1997; 97US-0047583.
XX 23-MAY-1997; 97US-0047584.
XX 23-MAY-1997; 97US-0047585.
XX 23-MAY-1997; 97US-0047586.
XX 23-MAY-1997; 97US-0047587.
XX 23-MAY-1997; 97US-0047588.
XX 23-MAY-1997; 97US-0047589.
XX 23-MAY-1997; 97US-0047590.
XX 23-MAY-1997; 97US-0047592.
XX 23-MAY-1997; 97US-0047593.
XX 23-MAY-1997; 97US-0047594.
XX 23-MAY-1997; 97US-0047595.
XX 23-MAY-1997; 97US-0047596.
XX 23-MAY-1997; 97US-0047597.
XX 23-MAY-1997; 97US-0047598.
XX 23-MAY-1997; 97US-0047599.
XX 23-MAY-1997; 97US-0047600.
XX 23-MAY-1997; 97US-0047601.
XX 23-MAY-1997; 97US-0047612.
XX 23-MAY-1997; 97US-0047613.
XX 23-MAY-1997; 97US-0047614.
XX 23-MAY-1997; 97US-0047615.
XX 23-MAY-1997; 97US-0047616.
XX 23-MAY-1997; 97US-0047617.
XX 23-MAY-1997; 97US-0047618.
XX 23-MAY-1997; 97US-0047632.
XX 23-MAY-1997; 97US-0047633.
XX 06-JUN-1997; 97US-0048964.
XX 06-JUN-1997; 97US-0048974.
XX 13-JUN-1997; 97US-0049610.
XX 08-JUL-1997; 97US-0051926.
XX 16-JUL-1997; 97US-0052874.
XX 18-AUG-1997; 97US-0055724.
XX 22-AUG-1997; 97US-0056630.
XX 22-AUG-1997; 97US-0056631.
XX 22-AUG-1997; 97US-0056632.
XX 22-AUG-1997; 97US-0056636.

PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057669.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
XX Feng P, Ferlie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
XX Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
XX Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX WPI: 1998-506364/43.
XX P-PSDB; AAW74792.
XX
XX New isolated human genes and the secreted polypeptide(s) they encode
XX - useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders
XX
XX Claim 1; Page 293-294; 721pp; English.
XX
XX This sequence represents a nucleic acid molecule designated Gene 62 from
XX the human cDNA clone HATDT67 (deposited as clone ATCC 97900 and ATCC
XX 209046) which encodes a secreted human protein. The gene can be used to
XX generate fusion proteins by linking to the gene to a human
XX immunoglobulin FC portion (e.g. AAV59502) for increasing the stability of
XX the fused protein as compared to the human protein only.
XX The invention relates to 186 novel genes and their fragments (nucleic
XX acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)
XX which are useful for preventing, treating or ameliorating medical
XX conditions e.g. by protein or gene therapy. Also, pathological conditions
XX can be diagnosed by determining the amount of the new polypeptides in a
XX sample or by determining the presence of mutations in the new
XX polynucleotides. Specific uses are described for each of the 186
XX polynucleotides, based on which tissues they are most highly expressed in
XX (see AAV59511 for described uses).
XX
XX Sequence 713 BP; 252 A; 115 C; 98 G; 248 T; 0 other;
XX
Query Match 7.2%; Score 19; DB 19; Length 713;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```
XX PA (EPiG-) EPIGENOMICS AG.
XX PT Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2001-602752/68.
XX XX
XX PT Fragments of chemically modified genes associated with tumour suppressor
XX PT genes and oncogenes, useful in designing primers and probes for
XX PT analysing diseases associated with cytosine methylation state e.g.
XX PT cancer
XX XX
XX PS Claim 1; SEQ ID No 470; 27pp; English.
XX XX
XX CC The invention relates to a nucleic acid comprising a sequence of 18
XX CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX CC bisulphite, of genes associated with tumour suppression and
XX CC oncogenes having a sequence taken from 536 (actually 533 since
XX CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
XX CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
XX CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
XX CC form part of a set of probes for detecting the cytosine methylation state
XX CC and/or single nucleotide polymorphisms and also to be used in an
XX CC array for analysing diseases associated with CpG dinucleotides e.g.
XX CC cancers and tumours. The probes can also be used in a method for
XX CC ascertaining genetic and/or epigenetic parameters for the diagnosis
XX CC and/or therapy of existing diseases or the predisposition to specific
XX CC diseases, by analysing cytosine methylations. The parameters may be
XX CC compared to another set of genetic and/or epigenetic parameters, the
XX CC differences serving as basis for diagnosis and/or prognosis events which
XX CC are disadvantageous to patients. The present sequence is one of the
XX CC 533 genomic sequences derived from tumour suppressor genes and
XX CC oncogenes. Sequences with even numbered Seq ID numbers are the
XX CC complementary sequences of the corresponding odd numbered sequence (e.g.
XX CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
XX CC is missing).
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX XX
XX SQ Sequence 38342 BP; 12471 A; 472 C; 7153 G; 18246 T; 0 other;

Query Match 7.5%; Score 20; DB 22; Length 38342;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 tttaggtatgtttgtttg 259
|||||
Db 9908 tttaggtatgtttgtttg 9927

RESULT 14
AAZ40793/c
ID AAZ40793 standard; DNA; 687 BP.
XX AC AAZ40793;
XX XX
XX DT 18-JAN-2000 (first entry)
XX XX
XX DE Secreted protein EST coding sequence 108-004-5-0-F5-FL.
XX XX
XX KW Secreted protein; fingerprint identification technique;
XX KW chromosome mapping; human; hereditary disease; diagnosis; cancer;
XX KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
XX KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
XX KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;
XX KW hypertension; ss.
XX OS Homo sapiens.
XX XX
XX PN W09940189-A2.
```

```
XX PD 12-AUG-1999.
XX XX
XX PF 09-FEB-1999; 99WO-IB00282.
XX XX
XX PR 09-FEB-1998; 98US-0074121.
XX PR 13-APR-1998; 98US-0081563.
XX PR 10-AUG-1998; 98US-0096116.
XX PR 04-SEP-1998; 98US-0099273.
XX PA (GEST ) GENSET.
XX XX
XX PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX XX
XX DR WPI; 1999-600966/51.
XX DR P-PSDB; RAY59665.
XX XX
XX PT Extended cDNAs useful for expressing secreted proteins and to obtain
XX PT specific antibodies -
XX XX
XX PS Claim 1; Page 159; 244pp; English.
XX XX
XX CC This sequence encodes a human secreted protein of the invention. The
XX CC extended cDNAs (or genomic DNA obtainable from them) may be used to
XX CC prepare PCR primers and probes. These are useful for forensic matching or
XX CC positive identification by DNA sequencing. They may also be used in
XX CC alternative fingerprint identification techniques. Antibodies against the
XX CC proteins encoded by the extended cDNAs are useful in identification of
XX CC tissue types or cell species, as well as identifying tissue specific
XX CC soluble proteins. The sequences can be used for chromosome mapping and
XX CC identification of genes associated with hereditary diseases or drug
XX CC response. Signal sequences from the cDNAs can be used in construction of
XX CC secretion vectors. Other sequences derived from the extended cDNAs can be
XX CC used to clone upstream genomic DNA sequences including promoters. This is
XX CC in turn useful for identifying proteins that interact with promoter
XX CC sequences. Some of the proteins may be useful in diagnosing and treating
XX CC several disorders including, but not limited to: cancer, hyperlipidaemia,
XX CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
XX CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,
XX CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.
XX XX
XX SQ Sequence 687 BP; 251 A; 126 C; 107 G; 203 T; 0 other;

Query Match 7.2%; Score 19; DB 20; Length 687;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 gtatgtttgtttgtttt 263
|||||
Db 641 GfATGTTTGTGTTGTTT 623

RESULT 15
AAV59572
ID AAV59572 standard; DNA; 713 BP.
XX AC AAV59572;
XX XX
XX DT 06-JAN-1999 (first entry)
XX XX
XX DE Human secreted protein gene 62 clone HATDT67.
XX XX
XX KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
XX KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
XX KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
XX KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
XX KW inflammation; ischaemic shock; Alzheimer's disease; testenosis; AIDS;
XX KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
XX KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
XX KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX XX
XX OS Homo sapiens.
```


PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-465570/50.

DR Isolated nucleic acid molecule encoding a reproductive system antigen -
PT is used in preventing, treating or ameliorating a medical condition -

XX Disclosure; SEQ ID NO 7363; 1297pp + Sequence Listing; English.

XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.

XX SQ Sequence 5862 BP; 1521 A; 1329 C; 1467 G; 1545 T; 0 other;

Query Match 7.5%; Score 20; DB 22; Length 5862;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 246 tatgtttgtttgtttgttc 265
Db 3760 tatgtttgtttgtttgttc 3779
RESULT 12
AAL04676
ID AAL04676 standard; DNA; 32249 BP.
XX AC AAL04676;
XX XX 21-NOV-2001 (first entry)
XX Human reproductive system related antigen DNA SEQ ID NO: 7364.
XX Human reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX Homo sapiens.
XX WO200155320-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01339.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.

PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases -
XX
PS Disclosure; SEQ ID No 2097; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides (AAU2915-AAU23814), and the cDNA and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAS41685-AA342192 represent DNA sequences encoding for the novel human
CC enzyme polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 3842 BP; 773 A; 1003 C; 1048 G; 1018 T; 0 other;

Query Match 7.58; Score 20; DB 22; Length 3842;

Best Local Similarity 100.0%; Pred. No. 3.1;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 aggtatgtttgtttgtttt 262

|||||

Db 358 aggtatgtttgtttgtttt 377

RESULT 11

AA04675

ID AAL04675 standard; DNA; 5862 BP.

XX AAL04675;

AC AAL04675;

XX 21-NOV-2001 (first entry)

DE Human reproductive system related antigen DNA SEQ ID NO: 7363.

XX Human; reproductive system related antigen; reproductive system disorder;

KW cancer; gene therapy; ds.

XX Homo sapiens.

XX WO200155320-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01339.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226888.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465566/50.
XX
XX Novel polypeptides and polynucleotides useful for diagnosing,
XX preventing, treating neural, immune system, muscular, reproductive,
PT

```
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 03-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.

PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465566/50.
XX Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases
XX
XX Disclosure: SEQ ID No 2096; 1180pp; English.
XX
XX The present invention relates to the isolation of novel human enzyme
CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAS41685-AAS42192 represent DNA sequences encoding for the novel human
CC enzyme polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3702 BP; 751 A; 959 C; 1020 G; 972 T; 0 other;

Query Match 7.5%; Score 20; DB 22; Length 3702;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 243 aggtatgtttgtttgttt 262
|||||
Db 358 aggtatgtttgtttgttt 377

RESULT 10
AAS41971
ID AAS41971 standard; DNA; 3842 BP.
XX AAS41971;
XX
XX 17-DEC-2001 (first entry)
XX
XX Genomic sequence #287 encoding novel human enzyme polypeptide:
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; ds.
XX
XX Homo sapiens.
OS
XX WO200155301-A2.
PN
XX 02-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US01239.
```

AC AAH94544;
 XX 05-OCT-2001 (first entry)
 XX Human foetal cDNA, SEQ ID NO: 1231.
 XX Human: foetal protein; cytostatic; immunosuppressive; immunostimulant;
 KW neutropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
 KW gene therapy; antisense therapy; cancer; immune disorder;
 KW growth disorder; osteoporosis; thrombolytic disorder;
 KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.
 XX
 XX Homo sapiens.
 OS
 XX WO200155339-A2.
 PN
 XX 02-AUG-2001.
 PD
 XX 25-JAN-2001; 2001WO-US02723.
 PF
 XX 25-JAN-2000; 2000US-0491404.
 XX 15-SEP-2000; 2000US-0663870.
 PR 06-NOV-2000; 2000US-0707351.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;
 PI Liu C, Asundi V, Zhou P, Werhman T;
 PI
 XX WPI; 2001-465571/50.
 DR P-PSDB; AAM08869.
 XX
 XX Novel fetal proteins useful for the treatment and diagnosis of diseases
 PT associated with dysfunction of the protein e.g. cancers, immune
 PT disorders, growth disorders, thrombolytic disorders, nervous system
 PT disorders and inflammation -
 PT
 XX Example 3; Page 695-696; 715pp; English.
 PS
 XX The invention relates to novel foetal polypeptides encoded by
 CC polynucleotides comprising one of 477 sequences fully defined in the
 CC specification. The foetal polynucleotides and polypeptides are
 CC useful in the treatment and diagnosis of diseases such as cancers,
 CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
 CC disorders, nervous system disorders and inflammation. The present
 CC sequence is a full length cDNA which was assembled using expressed
 CC sequence tags (ESTs) found to be expressed in human foetal tissue
 CC cDNA libraries as seeds.
 XX
 SQ Sequence 1158 BP; 341 A; 196 C; 244 G; 377 T; 0 other;

Query Match 7.5%; Score 20; DB 22; Length 1158;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 gtagtgggtttgtttgtttg 264
 |||||
 Db 497 gtagtgggtttgtttgtttg 516

RESULT 9
 AAS41970
 ID AAS41970 standard; DNA; 3702 BP.
 XX
 AC AAS41970;
 XX
 XX 17-DEC-2001 (first entry)
 DT
 XX Genomic sequence #286 encoding novel human enzyme polypeptide.
 DE
 XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KW ligase; hyperproliferative disorder; immunodeficiency disorder;

KW autoimmune disorder; neurological disorder; metabolic disorder;
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
 KW anti arthritic; nephrotropic; anticoagulant; ds.
 XX
 OS Homo sapiens.
 XX WO200155301-A2.
 PN
 XX 02-AUG-2001.
 PD
 XX 17-JAN-2001; 2001WO-US01239.
 PF
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226688.
 PR 23-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.

```
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140931.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145931.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146388.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 13-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
```

```
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 7.5%; Score 20; DB 21; Length 619;
Best Local Similarity 100.0%; Pred. No. 3.2;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 245 gatatgtttgtttgtttg 264
    |||||
Db 25 GTATGTTTGTTCGTTTG 6
```

RESULT 8
AAH94544
ID AAH94544 standard; cDNA; 1158 BP.
XX

PR 10-AUG-1999;	99US-01481171.	Db 25 GTATGTTTTGTTTCTTTTG 6
PR 11-AUG-1999;	99US-01483119.	
PR 12-AUG-1999;	99US-0148341.	RESULT 7
PR 13-AUG-1999;	99US-0148565.	AAC38598/c
PR 13-AUG-1999;	99US-0148684.	ID AAC38598 standard; DNA; 619 BP.
PR 16-AUG-1999;	99US-0149358.	XX AAC38598;
PR 17-AUG-1999;	99US-01491175.	XX AAC38598;
PR 18-AUG-1999;	99US-0149426.	XX 17-OCT-2000 (first entry)
PR 20-AUG-1999;	99US-0149722.	XX 17-OCT-2000 (first entry)
PR 20-AUG-1999;	99US-0149723.	XX Arabidopsis thaliana DNA fragment SEQ ID NO: 21554.
PR 20-AUG-1999;	99US-0149929.	XX Arabidopsis thaliana
PR 23-AUG-1999;	99US-0149902.	XX Arabidopsis thaliana
PR 23-AUG-1999;	99US-0149930.	XX Arabidopsis thaliana
PR 25-AUG-1999;	99US-0150566.	XX Arabidopsis thaliana
PR 26-AUG-1999;	99US-0150884.	XX Arabidopsis thaliana
PR 27-AUG-1999;	99US-0151065.	XX Arabidopsis thaliana
PR 27-AUG-1999;	99US-0151066.	XX Arabidopsis thaliana
PR 27-AUG-1999;	99US-0151080.	XX Arabidopsis thaliana
PR 30-AUG-1999;	99US-0151303.	XX Arabidopsis thaliana
PR 31-AUG-1999;	99US-0151438.	XX Arabidopsis thaliana
PR 01-SEP-1999;	99US-0151930.	XX Arabidopsis thaliana
PR 07-SEP-1999;	99US-0152363.	XX Arabidopsis thaliana
PR 10-SEP-1999;	99US-0153070.	XX Arabidopsis thaliana
PR 13-SEP-1999;	99US-0153758.	XX Arabidopsis thaliana
PR 15-SEP-1999;	99US-0154018.	XX Arabidopsis thaliana
PR 16-SEP-1999;	99US-0154039.	XX Arabidopsis thaliana
PR 20-SEP-1999;	99US-0154779.	XX Arabidopsis thaliana
PR 22-SEP-1999;	99US-0155139.	XX Arabidopsis thaliana
PR 23-SEP-1999;	99US-0155486.	XX Arabidopsis thaliana
PR 24-SEP-1999;	99US-0155659.	XX Arabidopsis thaliana
PR 28-SEP-1999;	99US-0156458.	XX Arabidopsis thaliana
PR 29-SEP-1999;	99US-0156596.	XX Arabidopsis thaliana
PR 04-OCT-1999;	99US-0157117.	XX Arabidopsis thaliana
PR 05-OCT-1999;	99US-0157753.	XX Arabidopsis thaliana
PR 06-OCT-1999;	99US-0157865.	XX Arabidopsis thaliana
PR 07-OCT-1999;	99US-0158029.	XX Arabidopsis thaliana
PR 08-OCT-1999;	99US-0158232.	XX Arabidopsis thaliana
PR 12-OCT-1999;	99US-0158369.	XX Arabidopsis thaliana
PR 13-OCT-1999;	99US-0159293.	XX Arabidopsis thaliana
PR 13-OCT-1999;	99US-0159294.	XX Arabidopsis thaliana
PR 14-OCT-1999;	99US-0159329.	XX Arabidopsis thaliana
PR 14-OCT-1999;	99US-0159330.	XX Arabidopsis thaliana
PR 14-OCT-1999;	99US-0159331.	XX Arabidopsis thaliana
PR 14-OCT-1999;	99US-0159637.	XX Arabidopsis thaliana
PR 18-OCT-1999;	99US-0159638.	XX Arabidopsis thaliana
PR 21-OCT-1999;	99US-0159584.	XX Arabidopsis thaliana
PR 21-OCT-1999;	99US-0160741.	XX Arabidopsis thaliana
PR 21-OCT-1999;	99US-0160767.	XX Arabidopsis thaliana
PR 21-OCT-1999;	99US-0160768.	XX Arabidopsis thaliana
PR 21-OCT-1999;	99US-0160770.	XX Arabidopsis thaliana
PR 21-OCT-1999;	99US-0160814.	XX Arabidopsis thaliana
PR 21-OCT-1999;	99US-0160815.	XX Arabidopsis thaliana
PR 22-OCT-1999;	99US-0160980.	XX Arabidopsis thaliana
PR 22-OCT-1999;	99US-0160981.	XX Arabidopsis thaliana
PR 22-OCT-1999;	99US-0160989.	XX Arabidopsis thaliana
PR 25-OCT-1999;	99US-0161404.	XX Arabidopsis thaliana
PR 25-OCT-1999;	99US-0161405.	XX Arabidopsis thaliana
PR 25-OCT-1999;	99US-0161406.	XX Arabidopsis thaliana
PR 26-OCT-1999;	99US-0161359.	XX Arabidopsis thaliana
PR 26-OCT-1999;	99US-0161360.	XX Arabidopsis thaliana
PR 26-OCT-1999;	99US-0161361.	XX Arabidopsis thaliana
PR 28-OCT-1999;	99US-0161920.	XX Arabidopsis thaliana
PR 28-OCT-1999;	99US-0161992.	XX Arabidopsis thaliana
PR 28-OCT-1999;	99US-0161993.	XX Arabidopsis thaliana
PR 29-OCT-1999;	99US-0162142.	XX Arabidopsis thaliana

Query Match
Best Local Similarity 100.0%; Score 20; DB 21; Length 617;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Pred. No. 3.2;

Qy 245 statgattttgtttttttttg 264
|||||

Query Match 7.9%; Score 21; DB 22; Length 8253;
Best Local Similarity 100.0%; Pred. No. 0.99;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 244 ggtatgtttgtttgtttgtttg 264
|||||
Db 7317 ggtatgtttgtttgtttgtttg 7337

RESULT
AAC45859/c
ID AAC45859 standard; DNA; 617 BP.
XX
AC AAC45859;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 48023.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
FN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138340.

PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.

```

Query Match      17.7%; Score 47; DB 22; Length 1725;
Best Local Similarity 100.0%; Pred. No. 2e-13;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ccggagagggtcacagtaagtgtgatgataaaatgggttcacagctgc 47
    |||||
Db 1679 ccggagagggtcacagtaagtgtgatgataaaatgggttcacagctgc 1725

RESULT 5
ID AAD06430
AC AAD06430 standard; DNA; 8253 BP.
XX
XX
XX
XX 10-AUG-2001 (first entry)
XX
XX Human Mcl-1 genomic DNA.
XX
XX Human; Mcl-1 gene regulatory element; Mcl-1s/deltaTM variant;
KW neuronal cell; tumour cell; apoptosis; therapy; cancer; psoriasis;
KW diabetic retinopathy; corneal graft neovascularisation;
KW neovascular glaucoma; epithelial condition; autoimmune disease;
KW rheumatoid arthritis; systemic lupus erythematosus;
KW neurodegenerative disease; ds.
XX
XX Homo sapiens.
XX
XX Key
FH Location/Qualifiers
FT misc_signal 1..1657
FT /tag= a
FT /note= "Mcl-1 gene regulatory region"
FT misc_signal 1495..1657
FT /tag= b
FT /note= "Mcl-1 gene regulatory element"
FT misc_signal 1495..1550
FT /tag= c
FT /note= "Mcl-1 gene regulatory element"
FT misc_signal 1495..1564
FT /tag= d
FT /note= "Mcl-1 gene regulatory element"
FT misc_signal 1495..1606
FT /tag= e
FT /note= "Mcl-1 gene regulatory element"
FT misc_signal 1513..1550
FT /tag= f
FT /note= "Mcl-1 gene regulatory element"
FT misc_signal 1513..1564
FT /tag= g
FT /note= "Mcl-1 gene regulatory element"
FT misc_signal 1513..1606
FT /tag= h
FT /note= "Mcl-1 gene regulatory element"
FT protein_bind 1529..1536
FT /tag= i
FT /bound_moiety= "transcription factor Ets"
FT protein_bind 1539..1545
FT /tag= j
FT /bound_moiety= "transcription factor Sp1"
FT repeat_region 1542..1555
FT /tag= k
FT /rpt_type= "DIRECT"
FT repeat_unit 1542..1548
FT /tag= l
FT misc_feature 1547..1553
FT /tag= m
FT /note= "Region having dyad symmetry"
FT misc_signal 1550..1657
FT /tag= n
FT /note= "Mcl-1 gene regulatory element"
FT protein_bind 1551..1560
FT /tag= o
FT /bound_moiety= "transcription factor SRE"

```

```

FT misc_feature 1557..1564
FT /tag= p
FT /note= "Region having dyad symmetry"
FT misc_signal 1606..1657
FT /tag= q
FT /note= "Mcl-1 gene regulatory element"
FT misc_signal 1645..1358
FT /tag= r
FT /note= "Corresponds to two adjacent initiator
sequences (YYAXT/AYY) at the transcription start site"
FT misc_signal 1657
FT /tag= s
FT /note= "Transcription start site"
FT 5'UTR 1667..1726
FT /tag= t
FT CDS 1727..3884
FT /tag= u
FT /product= "Mcl-1 protein"
FT exon 1727..2414
FT /tag= v
FT /number= "1"
FT intron 2415..2765
FT /tag= w
FT /number= "1"
FT exon 2766..3013
FT /tag= x
FT /number= "2"
FT intron 3014..3736
FT /tag= y
FT /number= "2"
FT exon 3768..3884
FT /tag= z
FT /number= "3"
FT 3'UTR 6704..8253
FT /tag= aa
XX WO200136594-A1.
XX
XX 25-MAY-2001.
PD
PF 14-JAN-2000; 2000WO-US00969.
XX
XX 16-NOV-1999; 99US-0166113.
XX (DART-) DARTMOUTH COLLEGE.
PA
PI Craig RW, Bingle CD, Whyte M;
XX
XX WPI; 2001-343812/36.
DR P-FSDB; AAE02462, AAE02463.
XX
XX Novel Mcl-1 gene regulatory elements, useful for modulating expression
of Mcl-1 polypeptide or its variant which regulate apoptosis in
neuronal or tumor cells -
XX
XX Claim 13; Fig 1A-8; 125pp; English.
PS
XX
XX The present invention relates to Mcl-1 gene regulatory elements and the
variant Mcl-1s/deltaTM. The anti-apoptotic Mcl-1 protein is encoded by
exons 1, 2 and 3. The pro-apoptotic Mcl-1s/deltaTM variant encoded by
exons 1 and 3 is obtained due to alternative mRNA splicing. The Mcl-1
gene regulatory element is useful for modulating the Mcl-1 gene
expression in a cell e.g., neuronal cell or tumour cell, such that
apoptosis of the cell is induced or cell viability is increased. The
Mcl-1 and its regulatory elements are used for treating pathological
conditions which include cancer, diabetic retinopathy, corneal graft
neovascularisation and neovascular glaucoma, epithelial conditions such
as psoriasis, autoimmune diseases like rheumatoid arthritis, systemic
lupus erythematosus, and neurodegenerative diseases. The present genomic
DNA sequence encodes Mcl-1 protein.
XX
XX Sequence 8253 BP; 2153 A; 1815 C; 1959 G; 2326 T; 0 other;

```

DT 16-JAN-2002 (first entry)
 XX Human cDNA encoding Six-Transmembrane Protein of Prostate 1, STMP1.
 DE
 XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
 KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
 KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
 KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
 KW liver cancer; lung cancer; cytostatic; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH 178..1650
 FT CDS /*tag= a
 FT /product= "STMP1"
 FT
 XX WO200172962-A2.
 PN
 XX 04-OCT-2001.
 PD
 XX 23-MAR-2001; 2001WO-US09410.
 XX
 XX 24-MAR-2000; 2000US-191929P.
 XX (SAAT/) SAATCIOGLU F.
 XX Saatioglu F;
 XX WPI; 2001-662926/76.
 DR P-PSDB; AAU10187.
 DR
 XX New polynucleotide for the diagnosis, prevention and treatment for
 PT prostate and testis disorders, particularly prostate cancer, comprises
 PT prostate-specific or testis-specific nucleic acids
 XX
 XX Claim 4; Fig 4E; 114pp; English.
 XX
 XX The invention relates to substantially pure prostate-specific or
 CC testis-specific polypeptides and the nucleic acids encoding them.
 CC Also included are vectors and host cells expressing the proteins, a
 CC transgenic animal expressing the protein, antibodies against the
 CC proteins, probes for detecting the nucleic acids, antisense molecules
 CC for the nucleic acids and methods of isolating modulators of the
 CC proteins. Compounds that modulate the prostate specific or testis
 CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,
 CC cryptorchidism, undescended, retractile, ascending or vanished
 CC testis. Other proliferative disorders for which the modulators may be
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The
 CC present sequence encodes a prostate specific protein, Six-Transmembrane
 CC Protein of Prostate 1, STMP1.
 XX
 XX Sequence 1680 BP; 467 A; 334 C; 373 G; 506 T; 0 other;
 SQ
 Query Match 20.4%; Score 54; DB 22; Length 1680;
 Best Local Similarity 100.0%; Pred. No. 7.5e-17;
 Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ccgagaggggtcacagtaagtgtatgataaagtgttcacagctgccatataa 54
 |||||
 Db 1627 ccgagaggggtcacagtaagtgtatgataaagtgttcacagctgccatataa 1680
 RESULT 4
 AAS15793
 ID AAS15793 standard; DNA: 1725 BP.
 XX
 AC AAS15793;
 XX

DT 16-JAN-2002 (first entry)
 XX Human DNA for Six-Transmembrane Protein of Prostate 1, STMP1.
 DE
 XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
 KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
 KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
 KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
 KW liver cancer; lung cancer; cytostatic; ds.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH 89
 FT misc_feature /*tag= a
 FT /note= "Represents 338 nucleotides of intron 1"
 FT
 FT 162
 FT misc_feature /*tag= b
 FT /note= "Represents 12713 nucleotides of intron 2"
 FT 200..1702
 FT CDS /*tag= c
 FT /product= "STMP1"
 FT 697
 FT misc_feature /*tag= d
 FT /note= "Represents 1396 nucleotides of intron 3"
 FT 1225
 FT misc_feature /*tag= e
 FT /note= "Represents 2372 nucleotides of intron 4"
 FT 1410
 FT misc_feature /*tag= f
 FT /note= "Represents 2299 nucleotides of intron 5"
 FT
 XX WO200172962-A2.
 PN
 XX 04-OCT-2001.
 PD
 XX 23-MAR-2001; 2001WO-US09410.
 XX
 XX 24-MAR-2000; 2000US-191929P.
 XX (SAAT/) SAATCIOGLU F.
 XX Saatioglu F;
 XX WPI; 2001-662926/76.
 DR P-PSDB; AAU10187.
 DR
 XX New polynucleotide for the diagnosis, prevention and treatment for
 PT prostate and testis disorders, particularly prostate cancer, comprises
 PT prostate-specific or testis-specific nucleic acids
 XX
 XX Example 3; Fig 4B; 114pp; English.
 XX
 XX The invention relates to substantially pure prostate-specific or
 CC testis-specific polypeptides and the nucleic acids encoding them.
 CC Also included are vectors and host cells expressing the proteins, a
 CC transgenic animal expressing the protein, antibodies against the
 CC proteins, probes for detecting the nucleic acids, antisense molecules
 CC for the nucleic acids and methods of isolating modulators of the
 CC proteins. Compounds that modulate the prostate specific or testis
 CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,
 CC cryptorchidism, undescended, retractile, ascending or vanished
 CC testis. Other proliferative disorders for which the modulators may be
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The
 CC present sequence encodes a prostate specific protein, Six-Transmembrane
 CC Protein of Prostate 1, STMP1.
 XX
 XX Sequence 1725 BP; 476 A; 340 C; 387 G; 517 T; 5 other;
 SQ

PT New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
XX prostate-specific or testis-specific nucleic acids -
PS Claim 6; Fig 4C; 114pp; English.
XX
CC The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence represents partial exon 6/3' UTR sequence of a prostate
CC specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.
XX
SQ Sequence 2381 BP; 780 A; 415 C; 387 G; 799 T; 0 other;

Query Match 100.0%; Score 265; DB 22; Length 2381;
Best Local Similarity 100.0%; Pred. No. 5.2e-120;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ccggagagggtcacagtaagtgtgatgataaaatgggttcacagctgccatataaagtctt 60
DB 117 ccggagagggtcacagtaagtgtgatgataaaatgggttcacagctgccatataaagtctt 176
QY 61 actcatgccattattttatgactcttcacgttcagttcacagtgctgtcacaattatcg 120
DB 177 actcatgccattattttatgactcttcacgttcagttcacagtgctgtcacaattatcg 236
QY 121 tgggttgaacctgttaaatgagatttcaactgacttagtgatagattttcttcaagtt 180
DB 237 tgggttgaacctgttaaatgagatttcaactgacttagtgatagattttcttcaagtt 296
QY 181 aattttcacaatgctcatgtttgccaatgatgaattttctagtcacacattatttgaat 240
DB 297 aattttcacaatgctcatgtttgccaatgatgaattttctagtcacacattatttgaat 356
QY 241 tttagtgattttttgtttgtttgc 265
DB 357 tttagtgattttttgtttgtttgc 381

RESULT 2
AAS15801
ID AAS15801 standard; cDNA; 4329 BP.
XX
AC AAS15801;
XX
DT 16-JAN-2002 (first entry)
XX
DE Human ORF of Six-Transmembrane Protein of Prostate 1, STMP1.
XX
KW Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 178..1650
FT CDS /*tag= a
FT /*product= "STMP1"

XX WO200172962-A2.
PN
XX
PD 04-OCT-2001.
XX
PF 23-MAR-2001; 2001WO-US09410.
XX
PR 24-MAR-2000; 2000US-191929P.
PA (SAAT/) SAATCIOGLU F.
XX
PI Saatcioglu F;
XX
DR WPI; 2001-662926/76.
DR P-PSDB; AAU10187.
XX
PT New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids -
PS Claim 5; Fig 4D; 114pp; English.
XX
CC The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence represents the open reading frame of a prostate
CC specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.
XX
SQ Sequence 4329 BP; 1315 A; 817 C; 790 G; 1407 T; 0 other;

Query Match 100.0%; Score 265; DB 22; Length 4329;
Best Local Similarity 100.0%; Pred. No. 5.1e-120;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ccggagagggtcacagtaagtgtgatgataaaatgggttcacagctgccatataaagtctt 60
DB 1627 ccggagagggtcacagtaagtgtgatgataaaatgggttcacagctgccatataaagtctt 1686
QY 61 actcatgccattattttatgactcttcacgttcagttcacagtgctgtcacaattatcg 120
DB 1687 actcatgccattattttatgactcttcacgttcagttcacagtgctgtcacaattatcg 1746
QY 121 tgggttgaacctgttaaatgagatttcaactgacttagtgatagattttcttcaagtt 180
DB 1747 tgggttgaacctgttaaatgagatttcaactgacttagtgatagattttcttcaagtt 1806
QY 181 aattttcacaatgctcatgtttgccaatgatgaattttctagtcacacattatttgaat 240
DB 1807 aattttcacaatgctcatgtttgccaatgatgaattttctagtcacacattatttgaat 1866
QY 241 tttagtgattttttgtttgtttgc 265
DB 1867 tttagtgattttttgtttgtttgc 1891

RESULT 3
AAS15802
ID AAS15802 standard; cDNA; 1680 BP.
XX
AC AAS15802;
XX

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:59:59 ; Search time 600.75 Seconds
(without alignments)
757.357 Million cell updates/sec

Title: US-09-802-520-9

Perfect score: 265

Sequence: 1 ccggagaggggtcacagtaat.....tatgtttgtttgtttgttc 265

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_032802.*

- 1: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1980.DAT.*
- 2: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1981.DAT.*
- 3: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1982.DAT.*
- 4: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1983.DAT.*
- 5: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1984.DAT.*
- 6: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1985.DAT.*
- 7: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1986.DAT.*
- 8: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1987.DAT.*
- 9: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1988.DAT.*
- 10: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1989.DAT.*
- 11: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1990.DAT.*
- 12: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1991.DAT.*
- 13: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1992.DAT.*
- 14: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1993.DAT.*
- 15: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1994.DAT.*
- 16: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1995.DAT.*
- 17: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1996.DAT.*
- 18: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.*
- 19: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1998.DAT.*
- 20: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1999.DAT.*
- 21: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2000.DAT.*
- 22: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001A.DAT.*
- 23: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001B.DAT.*
- 24: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	265	100.0	2381	22	AA515800 Human Six-Transmem
2	265	100.0	4329	22	AA515801 Human ORF of Six-T
3	54	20.4	1680	22	AA515802 Human cDNA encodin
4	47	17.7	1725	22	AA515793 Human DNA for Six-
5	21	7.9	8253	22	AAD06430 Human Mcl-1 genomi
6	20	7.5	617	21	AA45859 Arabidopsis thalia
7	20	7.5	619	21	AA45859 Arabidopsis thalia
8	20	7.5	1158	22	AAH94544 Human foetal cDNA,
9	20	7.5	3702	22	AA541970 Genomic sequence #

10	20	7.5	3842	22	AA541971 Genomic sequence #
11	20	7.5	5862	22	AA541971 Human reproductive
12	20	7.5	32249	22	AA541971 Human reproductive
13	20	7.5	38342	22	AA541971 Tumour suppressor
14	19	7.2	687	20	AA541971 Secreted protein E
15	19	7.2	713	19	AA541971 Human secreted pro
16	19	7.2	830	19	AA541971 Human secreted pro
17	19	7.2	1234	9	AA541971 Sequence encoding
18	19	7.2	1980	22	AA541971 Human sub-unit C o
19	19	7.2	2540	23	AA541971 Drosophila melanog
20	19	7.2	2938	23	AA541971 Drosophila melanog
21	19	7.2	4544	22	AA541971 Human lung antigen
22	19	7.2	4706	23	AA541971 Drosophila melanog
23	19	7.2	8724	22	AA541971 Human immune/haema
24	19	7.2	9963	24	AA541971 Human immune syste
25	19	7.2	16579	22	AA541971 Human nervous syst
26	19	7.2	16579	22	AA541971 Human reproductive
27	19	7.2	25402	22	AA541971 Human immune/haema
28	18	6.8	147	16	AA541971 Human gene signatu
29	18	6.8	155	24	AA541971 Activated T-cell d
30	18	6.8	217	16	AA541971 Human gene signatu
31	18	6.8	251	22	AA541971 Human immune/haema
32	18	6.8	252	18	AA541971 Murine metastatic
33	18	6.8	254	21	AA541971 Human secreted pro
34	18	6.8	255	22	AA541971 Rat differential t
35	18	6.8	259	20	AA541971 EST clone GE89. H
36	18	6.8	300	20	AA541971 Human gene express
37	18	6.8	300	20	AA541971 Human gene express
38	18	6.8	317	20	AA541971 Human DNA marker c
39	18	6.8	337	22	AA541971 Human breast cance
40	18	6.8	375	22	AA541971 Novel human diagno
41	18	6.8	414	22	AA541971 Human polynucleoti
42	18	6.8	432	22	AA541971 Human immune/haema
43	18	6.8	432	22	AA541971 Human immune/haema
44	18	6.8	444	22	AA541971 Human polynucleoti
45	18	6.8	455	22	AA541971 Human immune/haema

ALIGNMENTS

RESULT 1	
AA515800	AA515800 standard; DNA; 2381 BP.
ID	AA515800
AC	AA515800;
XX	
DT	16-JAN-2002 (first entry)
XX	
DE	Human Six-Transmembrane Protein of Prostate 1, STMP1, exon 6/3' UTR.
XX	
KW	Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW	benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW	cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW	leukemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW	liver cancer; lung cancer; cytostatic; ds; exon 6.
OS	Homo sapiens.
XX	
PN	WO200172962-A2.
XX	
PD	04-OCT-2001.
XX	
PF	23-MAR-2001; 2001WO-US09410.
XX	
PR	24-MAR-2000; 2000US-191929P.
XX	
PA	(SAAT/) SAATCIOGLU F.
XX	
PI	Saatcioglu F;
XX	
DR	WPI; 2001-662926/76.
XX	

match: proteins: Tr:Q9P0F6"
/codon_start=1

Query Match
Best Local Similarity 7.9%; Score 21; DB 9; Length 135005;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 ggtatgttttggttgtttg 264
|||||
Db 38364 GGTATGTTTGTGTTGTTG 38384

misc_feature 130263..140554
/note="assembly_fragment:03416.0"
BASE COUNT 40051 a 30196 c 30768 g 39134 t 405 others
ORIGIN

Query Match 7.9%; Score 21; DB 2; Length 140554;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 ggtatgttttggttgtttg 264
|||||
Db 12877 GGTATGTTTGTGTTGTTG 12897

Search completed: September 20, 2002, 09:59:39
Job time: 13437 sec

AL645757 140554 bp DNA linear HTG 08-FEB-2002
Mus musculus chromosome 3 clone XxBac-340J1, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.

ACCESSION AL645757
VERSION AL645757.10 GI:18643801
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
Clark,G.
Direct Submission
Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 10, 2002 this sequence version replaced gi:17384176.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bq340J1
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 139891 bases at least Q40
Consensus quality: 139999 bases at least Q30
Consensus quality: 140052 bases at least Q20
Insert size: 140154; sum-of-contigs
Quality coverage: 142955; 2.7% error; agarose-fp
Quality coverage: 12.79x in Q20 bases; sum-of-contigs Quality
coverage: 12.58x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source
1..140554
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="3"
/clone="XxBac-340J1"
/clone_lib="NOD mouse library"
1..40212
/note="assembly_fragment:00914
fragment_chain:1"
40313..46487
/note="assembly_fragment:04023
fragment_chain:1"
46588..49183
/note="assembly_fragment:03541
fragment_chain:2"
49284..130162
/note="assembly_fragment:04031
fragment_chain:2"


```
source
1..109395
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP11-115O13"
/clone_lib="RPC1-11 Human Male BAC"
1..1076
/note="assembly_fragment"
vector_end:SP6
vector_side:left"
1177..2259
/note="assembly_fragment"
2360..3476
/note="assembly_fragment"
3577..4665
/note="assembly_fragment"
4766..5957
/note="assembly_fragment"
6058..7087
/note="assembly_fragment"
7188..8243
/note="assembly_fragment"
8344..9433
/note="assembly_fragment"
9534..10942
/note="assembly_fragment"
11043..12544
/note="assembly_fragment"
12645..13988
/note="assembly_fragment"
14089..15211
/note="assembly_fragment"
15312..16697
/note="assembly_fragment"
16798..18219
/note="assembly_fragment"
18320..19469
/note="assembly_fragment"
19570..21074
/note="assembly_fragment"
21175..22647
/note="assembly_fragment"
22748..24713
/note="assembly_fragment"
24814..26482
/note="assembly_fragment"
26583..28072
/note="assembly_fragment"
28173..29622
/note="assembly_fragment"
29723..32160
/note="assembly_fragment"
32261..34500
/note="assembly_fragment"
34601..36536
/note="assembly_fragment"
```

Query Match 7.9%; Score 21; DB 2; Length 109395;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 gtagtattttgtttgtttg 264
|||||

Db 98099 GGTATGTTTGTGTTTGTGTTG 98119

RESULT 13
AL390202_06
WPCOMMENT

Sequence split into 10 fragments LOCUS AL390202 Accession AL390202
Fragment Name Begin End
AL390202_00 1 110000
AL390202_01 100001 210000
AL390202_02 200001 310000

AL390202_03 300001 410000
AL390202_04 400001 510000
AL390202_05 500001 610000
AL390202_06 600001 710000
AL390202_07 700001 810000
AL390202_08 800001 910000
AL390202_09 900001 988176

Continuation (7 of 10) of AL390202 from base 600001 (AL390202 Homo sapiens chromosome

Query Match 7.9%; Score 21; DB 2; Length 110000;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 gtagtattttgtttgtttgc 265
|||||

Db 23381 GTATGTTTGTGTTTGTGTTGC 23401

RESULT 14

LOCUS

HS860F19

135005 bp DNA linear PRI 20-APR-2001

Human DNA sequence from clone RP5-860F19 on chromosome 20p12.3-13
Contains the gene for KIAA1442 (similar to olfactory neuronal
transcription factors (COE1, COE2, COE3, EBF3, OLF1)), RPL19 (60S
ribosomal protein L19) and HSPC080 pseudogenes, the gene for
metallocarboxypeptidase (CPX-1) and a novel gene. Contains ESTs,
STSs, GSSs and four CpG islands, complete sequence.

ACCESSION

AL035460

AL035460.15 GI:7321166

VERSION

HTG: COE; CpG island: CPX-1; EBF3; KIAA1442;

KEYWORDS

metallocarboxypeptidase; OLF1; RPL19.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS

Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.

JOURNAL

1 (bases 1 to 135005)

Direct Submission

Smith, W.

COMMENT

Submitted (20-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Mar 23, 2000 this sequence version replaced gi:6468345.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep

This sequence

was generated from part of bacterial clone contigs of human

chromosome 20, constructed by the Sanger Centre Chromosome 20

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr20

This sequence is the entire insert of clone RP5-860F19. The true

right end of clone RP5-1056J6 is at 4586 in this sequence. This

sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. RP5-860F19 is from

the library RPC1-5 constructed by the group of Pieter de Jong. For

further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pCYPAC2.

Location/Qualifiers

1..135005

FEATURES

source


```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="leukocyte"
/notes="isolated from a human leukocyte genomic library in
EMBL-3 from Clontech (HL 10006d)"
Join(1657. .2414,3768. .6703)
/gene="MCL-1"
/product="Mcl-1 delta S/TM"
Join(1657. .2414,2766. .3013,3768. .6703)
/gene="MCL-1"
/product="Mcl-1"
1657. .6703
/gene="MCL-1"
Join(1727. .2414,3768. .3895)
/gene="MCL-1"
/notes="alternatively spliced isoform of Mcl-1 has a
pro-apoptotic function."
/codon_start=1
/product="Mcl-1 delta S/TM"
/protein_id="AAF64256.1"
/db_xref="GI:7582272"
/translation="MFLGRNVAIGLNLYCGAGLGGGATRPGGRLLATEKEASA
RREIGGAGAVIGGSAGASPPSTLTPDSRRVAPPPIGAEVDPDVATPARLLFPAPT
RAAPLEMEAPADAINSPREELDGYEPLGKRPAYLPILLEVGESNNTSTDGSL
PSTPPAEEEDDLYRQSLIISLYLREQATGAKDTKPMGRSGATSRKALFTLRVGD
GVQRNHETAFQGVCGVLPGRPRWHQECAGFCRCWRSWFGISNKNALL"
Join(1727. .2414,2766. .3013,3768. .3884)
/gene="MCL-1"
/notes="anti-apoptotic Bcl-2 family member"
/codon_start=1
/product="Mcl-1"
/protein_id="AAF64255.1"
/db_xref="GI:7582271"
/translation="MFLGRNVAIGLNLYCGAGLGGGATRPGGRLLATEKEASA
RREIGGAGAVIGGSAGASPPSTLTPDSRRVAPPPIGAEVDPDVATPARLLFPAPT
RAAPLEMEAPADAINSPREELDGYEPLGKRPAYLPILLEVGESNNTSTDGSL
PSTPPAEEEDDLYRQSLIISLYLREQATGAKDTKPMGRSGATSRKALFTLRVGD
GVQRNHETAFQGVCGVLPGRPRWHQECAGFCRCWRSWFGISNKNALL"
FAGVAGVAGLAYLIR"
BASE COUNT 2153 a 1818 c 1958 g 2324 t
ORIGIN

Query Match 7.9%; Score 21; DB 9; Length 8253;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 gtagtattttgtttgtttg 264
|||||
Db 7317 GGTATGTTTGTGTTGTTG 7337

RESULT 11
AL592167/c LOCUS 10917 bp DNA linear PRI 09-AUG-2001
DEFINITION Human DNA sequence from clone RP11-14H2 on chromosome 13, complete
sequence.
ACCESSION AL592167
VERSION 1
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 10917)
AUTHORS Corby,N.
TITLE Direct Submission
SUBMITTED (09-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 10, 2001 this sequence version replaced gi:14626209.
During sequence assembly data is compared from overlapping clones.
```

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr13>

RP11-14H2 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-14H2. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-181D10 is at 8918 in this sequence.

The true right end of clone RP11-89L15 is at 2000 in this sequence.

Location/Qualifiers

1. .10917

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="13"

/clone="RP11-14H2"

/clone_lib="RPCI-11.1"

80. .177

/note="MIR repeat: matches 47. .161 of consensus"

repeat_region 2301. .2424

/note="MERSA repeat: matches 3. .125 of consensus"

repeat_region 2703. .2824

/note="61 copies 2 mer aa 61% conserved"

repeat_region 2931. .3016

/note="MIR repeat: matches 151. .233 of consensus"

repeat_region 3021. .3111

/note="L1ME3 repeat: matches 6060. .6153 of consensus"

repeat_region 3160. .3216

/note="L1MD2 repeat: matches 5872. .5928 of consensus"

repeat_region 3315. .3588

/note="L2 repeat: matches 2020. .2314 of consensus"

repeat_region 4217. .4276

/note="2 copies 30 mer 91% conserved"

repeat_region 9892. .10206

/note="AluX repeat: matches 1. .312 of consensus"

repeat_region 10450. .10585

/note="L2 repeat: matches 2158. .2307 of consensus"

repeat_region 10793. .10894

/note="L2 repeat: matches 2563. .2677 of consensus"

BASE COUNT 3452 a 2365 c 1954 g 3146 t

ORIGIN

Query Match 7.9%; Score 21; DB 9; Length 10917;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 gtagtattttgtttgtttg 265
|||||

Db 6457 GTATGTTTGTGTTGTTGC 6437

RESULT 12

```

* 19057 19156: gap of 100 bp
* 19157 23097: contig of 3941 bp in length
* 23098 23197: gap of 100 bp
* 23198 26423: contig of 3226 bp in length
* 26424 26523: gap of 100 bp
* 26524 29626: contig of 3103 bp in length
* 29627 29726: gap of 100 bp
* 29727 34097: contig of 4371 bp in length
* 34098 34197: gap of 100 bp
* 34198 37833: contig of 3636 bp in length
* 37834 37933: gap of 100 bp
* 37934 43795: contig of 5862 bp in length
* 43796 43895: gap of 100 bp
* 43896 48837: contig of 4942 bp in length
* 48838 48937: gap of 100 bp
* 48938 54368: contig of 5431 bp in length
* 54369 54468: gap of 100 bp
* 54469 61661: contig of 7193 bp in length
* 61662 61661: gap of 100 bp
* 61762 72335: contig of 10574 bp in length
* 72336 72435: gap of 100 bp
* 72436 81162: contig of 8727 bp in length
* 81163 81262: gap of 100 bp
* 81263 89194: contig of 7932 bp in length
* 89195 89294: gap of 100 bp
* 89295 99216: contig of 9922 bp in length
* 99217 99316: gap of 100 bp
* 99317 107444: contig of 8128 bp in length
* 107445 107544: gap of 100 bp
* 107545 118282: contig of 10738 bp in length
* 118283 118382: gap of 100 bp
* 118383 130386: contig of 12004 bp in length
* 130387 130486: gap of 100 bp
* 130487 142983: contig of 12497 bp in length
* 142984 143083: gap of 100 bp
* 143084 159402: contig of 16319 bp in length
* 159403 159502: gap of 100 bp
* 159503 179408: contig of 19906 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..179408
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /map="8"
                /clone="RP11-311H16"
                /clone_lib="RP11 Human Male BAC"
            1..1105
                /note="assembly_fragment"
            1206..2470
                /note="assembly_fragment"
            2571..4687
                /note="assembly_fragment"
            4788..6276
                /note="assembly_fragment"
            6377..7590
                /note="assembly_fragment"
            7691..8804
                /note="assembly_fragment"
            clone_end:T7
                vector_side:right
            8905..12597
                /note="assembly_fragment"
            12698..15143
                /note="assembly_fragment"
            15244..19056
                /note="assembly_fragment"
            19157..23097
                /note="assembly_fragment"
            23198..26423
                /note="assembly_fragment"
            26524..29626
                /note="assembly_fragment"
            29727..34097

```

```

misc_feature
    /note="assembly_fragment"
    34198..37833
misc_feature
    /note="assembly_fragment"
    37934..43795
misc_feature
    /note="assembly_fragment"
    43896..48837
misc_feature
    /note="assembly_fragment"
    48938..54368
misc_feature
    /note="assembly_fragment"
    54469..61661
misc_feature
    /note="assembly_fragment"
    61762..72335
misc_feature
    /note="assembly_fragment"
    72436..81162
misc_feature
    /note="assembly_fragment"
    81263..89194
misc_feature
    /note="assembly_fragment"
    89295..99216
misc_feature
    /note="assembly_fragment"
    99317..107444
misc_feature
    /note="assembly_fragment"
    clone_end:SP6
    vector_side:right
    107545..118282
misc_feature
    /note="assembly_fragment"
    118383..130386
misc_feature
    /note="assembly_fragment"
    130487..142983
misc_feature
    /note="assembly_fragment"
    143084..159402
misc_feature
    /note="assembly_fragment"
    159503..179408
misc_feature
    /note="assembly_fragment"
    179409..36807
BASE COUNT 52785 a 36085 c 36807 g 51028 t 2703 others
ORIGIN

```

```

Query Match      8.3%  Score 22;  DB 2;  Length 179408;
Best Local Similarity 100.0%;  Pred. No. 0.53;
Matches 22;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy 243 aggtatgtttgtttgtttgtttg 264
      |||||||||||||||||||
Db 149091 AGGTATGTTTGTGTTTGTGTTG 149112

```

RESULT 10

```

AF198614
LOCUS      AF198614                8253 bp    DNA    linear    PRI 17-JUL-2000
DEFINITION Homo sapiens Mcl-1 (MCL-1) and Mcl-1 delta S/TM (MCL-1) genes,
            alternative spliced forms, complete cds.
ACCESSION  AF198614
VERSION     AF198614.1  GI:7582270
KEYWORDS    human.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 8253)
AUTHORS     Bingle, C.D., Craig, R.W., Swales, B.M., Singleton, V., Zhou, P. and
            Whyte, M.K.
TITLE       Exon skipping in mcl-1 results in a bcl-2 homology domain 3 only
            gene product that promotes cell death
JOURNAL     J. Biol. Chem. 275 (29), 22136-22146 (2000)
MEDLINE     20357335
REFERENCE   2 (bases 1 to 8253)
AUTHORS     Craig, R.W., Zhou, P. and Bingle, C.D.
TITLE       Direct Submission
JOURNAL     Submitted (26-OCT-1999) Molecular and Genetic Medicine, University
            of Sheffield Medical School, Glossop Rd, Sheffield, S10
            2RX, UK
FEATURES    Location/Qualifiers
            source
                1..8253

```

Insert size: 170983; sum-of-contigs
Quality coverage.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1324: contig of 1324 bp in length
* 1325 1424: gap of 100 bp
* 1425 2513: contig of 1089 bp in length
* 2514 2613: gap of 100 bp
* 2614 23113: contig of 20500 bp in length
* 23114 23213: gap of 100 bp
* 23214 51238: contig of 28025 bp in length
* 51239 51338: gap of 100 bp
* 51339 81239: contig of 29901 bp in length
* 81240 81339: gap of 100 bp
* 81340 171483: contig of 90144 bp in length.

FEATURES

source
1. .171483
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-44J4"
/clone_lib="RPC1-11 Human Male BAC"
1. .1324
/note="assembly_fragment"
1425. .2513
/note="assembly_fragment"
2614. .23113
/note="assembly_fragment"
23214. .51238
/note="assembly_fragment"
51339. .81239
/note="assembly_fragment"
clone_end:SP6
vector_side:left
81340. .171483
/note="assembly_fragment"
clone_end:T7
vector_side:left
48824 a 33837 c 35795 g 52521 t 506 others
ORIGIN

Query Match 8.3%; Score 22; DB 2; Length 171483;
Best Local Similarity 100.0%; Pred. No. 0.53;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 aggtatgtttgtttgtttgtttg 264
|||||||
Db 43705 AGGTATGTTTGTGTTGTTTG 43684

RESULT 9

AC027234
LOCUS AC027234 179408 bp DNA linear HTG 02-MAY-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-311H16 map 8, WORKING DRAFT
SEQUENCE, 28 unordered pieces.
ACCESSION AC027234
VERSION AC027234.2 GI:7677908
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179408)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-311H16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179408)

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collumore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliou,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
JOURNAL

COMMENT

Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 2, 2000 this sequence version replaced gi:7331604.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genom Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8711

Center clone name: 311_H16

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 165594 bases at least Q40

Consensus quality: 172613 bases at least Q30

Consensus quality: 175438 bases at least Q20

Insert size: 185000; agarose-1p

Insert size: 176708; sum-of-contigs

Quality coverage: 3.9 in Q20 bases; agarose-1p

Quality coverage: 4.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1105: contig of 1105 bp in length

* 1106 1205: gap of 100 bp

* 1206 2470: contig of 1265 bp in length

* 2471 2570: gap of 100 bp

* 2571 4687: contig of 2117 bp in length

* 4688 4787: gap of 100 bp

* 4788 6276: contig of 1489 bp in length

* 6277 6376: gap of 100 bp

* 6377 7590: contig of 1214 bp in length

* 7591 7690: gap of 100 bp

* 7691 8804: contig of 1114 bp in length

* 8805 8904: gap of 100 bp

* 8905 12597: contig of 3693 bp in length

* 12598 12697: gap of 100 bp

* 12698 15143: contig of 2446 bp in length

* 15144 15243: gap of 100 bp

* 15244 19056: contig of 3813 bp in length

```
repeat_region /note="MER2 repeat: matches 283. .345 of consensus"
18615. .18795
repeat_region /note="MIR repeat: matches 3. .202 of consensus"
19450. .19503
repeat_region /note="L2 repeat: matches 2650. .2706 of consensus"
19668. .20448
repeat_region /note="L2 repeat: matches 1944. .2708 of consensus"
20504. .20794
repeat_region /note="AluSx repeat: matches 23. .312 of consensus"
20795. .20966
repeat_region /note="L2 repeat: matches 2272. .2466 of consensus"
21653. .21720
repeat_region /note="MIR repeat: matches 76. .144 of consensus"
22071. .22266
repeat_region /note="MIR repeat: matches 20. .214 of consensus"
22380. .22463
repeat_region /note="MIR repeat: matches 35. .122 of consensus"
23842. .24153
repeat_region /note="AluSx repeat: matches 1. .308 of consensus"
25349. .25588
repeat_region /note="L1 repeat: matches 3686. .3939 of consensus"
25653. .25956
repeat_region /note="AluJo repeat: matches 1. .301 of consensus"
26369. .26677
repeat_region /note="AluSx repeat: matches 1. .309 of consensus"
26937. .27286
repeat_region /note="MLT1A2 repeat: matches 11. .374 of consensus"
27408. .27641
repeat_region /note="MIR repeat: matches 15. .261 of consensus"
28885. .29071
repeat_region /note="MER5A repeat: matches 1. .189 of consensus"
31232. .31277
repeat_region /note="L2 repeat: matches 2661. .2705 of consensus"
31619. .31685
repeat_region /note="L2 repeat: matches 2356. .2424 of consensus"
31926. .32125
repeat_region /note="MIR repeat: matches 57. .254 of consensus"
33172. .33364
repeat_region /note="MIR repeat: matches 64. .259 of consensus"
33414. .33824
repeat_region /note="L1MB5 repeat: matches 5740. .6174 of consensus"
34111. .34142
repeat_region /note="MIR repeat: matches 117. .148 of consensus"
34443. .36275
repeat_region /note="L1MC4 repeat: matches 6207. .7977 of consensus"
37848. .37905
repeat_region /note="MIR repeat: matches 120. .180 of consensus"
38347. .38455
repeat_region /note="MIR repeat: matches 48. .148 of consensus"
38602. .38731
repeat_region /note="L2 repeat: matches 2570. .2695 of consensus"
38754. .38879
repeat_region /note="L2 repeat: matches 2575. .2706 of consensus"
41096. .41180
repeat_region /note="MER5B repeat: matches 62. .160 of consensus"
41181. .41349
repeat_region /note="FRAM repeat: matches 1. .166 of consensus"
41350. .41413
repeat_region /note="MER5B repeat: matches 3. .62 of consensus"
41446. .41601
repeat_region /note="MIR repeat: matches 70. .241 of consensus"
41603. .41661
repeat_region /note="MLT1A1 repeat: matches 2. .58 of consensus"
41818. .41904
repeat_region /note="MLT1D repeat: matches 2. .87 of consensus"
42612. .43013
repeat_region /note="MSTA repeat: matches 1. .426 of consensus"
43071. .43269
repeat_region /note="MIR repeat: matches 47. .256 of consensus"
43631. .44125
repeat_region /note="LTR33 repeat: matches 6. .509 of consensus"
44270. .44579
repeat_region /note="AluY repeat: matches 1. .310 of consensus"
```

```
repeat_region 44597. .44830
/note="MIR repeat: matches 2. .243 of consensus"
repeat_region 45096. .45117
/note="11 copies 2 mer tt 100% conserved"
misc_feature 46347. .46371
/note="Sequence from Clone PCR only. Sequence from
Query Match 8.3% Score 22; DB 9; Length 64855;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 243 aggtatgtttgtttgtttgtttg 264
|||||
Db 25646 AGGTATGTTTGTGTTGTTGTTG 25667

RESULT 8
AC009278/c
LOCUS
DEFINITION Homo sapiens clone RP11-44J4, WORKING DRAFT SEQUENCE, 6 unordered
pieces.
AC009278
AC009278.4 GI:8072422
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 171483)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,N., Lehoczy,J., Lied,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (12-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7248943.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1074
Center clone name: 44_J_4
----- Summary Statistics
Sequencing vector: M13; M77815; 64% of reads
Sequencing vector: Plasmid; n/a; 0.0% of reads
Chemistry: Dye-terminator Big Dye; 96% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165349 bases at least Q40
Consensus quality: 168407 bases at least Q30
Consensus quality: 169725 bases at least Q20
Insert size: 173000; agarose-fp
```

LPSEAYRDPCLSPASSLSRSCNSESSVESNYSYPYASPTQSPWQSPCVKPTTD
PEEGFPLSGCHLGSPRHSPTSPASITEESWLGARGSRPTSPCKRYSLSNGRQ
PSCSPHSPSTSPHSPSVSTEDTWLNTQYTSSAIVAAINALTTDSTLDLGGVP
IKSRKTALEHAPSVALKVEPAGEDLTPTPTSDPPEYVTOHLKKGAFCCOYLSVPO
ASYOWAKPSLPTSPYSMSPLALDWLPSPHSGPYELRVEQPKSHHRAHVETESRG
AVKASAGHPVIVOLHGYLENPLTLQLFEGTADDBLLRPHAFYQVHRITGKVTSTSH
EILISNTKYLEIPLPENNRRAIDCAGILKLRNSDIELRKGETDGRKTRRVLFR
VHIPOPNRTLSQVANSPIECORSAGELPLVEKQSDSYDVPIGKKRWLSGHNFLO
DSKVTFEKPDPGHVMEAKTRDLCKPNSLVVEIPFRNQRITSPAQVSFYVCNG
KKRSQYQRYTLPANGNSVFLTSSSELRGFFY"

BASE COUNT 809 a 1009 c 828 g 823 t
ORIGIN

Query Match 8.3%; Score 22; DB 10; Length 3469;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 243 aggtatgtttttttttttg 264
|||||
Db 2218 AGGTATGTTTGTGTTTGT 2197

RESULT 7

AL139219 64855 bp DNA linear PRI 11-APR-2001
LOCUS Human DNA sequence from clone RP5-1117P19 on chromosome
DEFINITION lp31.2-32.3, complete sequence.

ACCESSION

AL139219

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 64855)

Tromans, A.

Direct Submission

Submitted (11-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Apr 12, 2001 this sequence version replaced gi:13445305.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em.; EMBL; Sw.;

SWISSPROT; Tr.; TrEMBL; Wp.; WORMPEP; Information on the WORMPEP

database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping

Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr1

RP5-1117P19 is from the library RP5-1117P19 constructed by the group of

Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

FEATURES

Source

Location/Qualifiers

1..64855

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/map="p31.2-32.3"

/clone="RP5-1117P19"

/clone_lib="RPC1-5"

17..134

/note="L2 repeat: matches 2592. .2698 of consensus"

216..279

/note="L2 repeat: matches 2647. .2709 of consensus"

779..872

/note="MIR repeat: matches 46. .153 of consensus"

1616..1703

/note="MIR repeat: matches 66. .155 of consensus"

1704..2002

/note="AluX repeat: matches 1. .295 of consensus"

2003..2100

/note="MIR repeat: matches 155. .247 of consensus"

2394..2525

/note="MIR repeat: matches 5. .140 of consensus"

2789..3020

/note="L2 repeat: matches 2511. .2745 of consensus"

3106..3228

/note="MIR repeat: matches 71. .191 of consensus"

3508..3572

/note="MIR repeat: matches 158. .219 of consensus"

4870..5108

/note="L2 repeat: matches 2248. .2500 of consensus"

5551..5981

/note="MLTIC repeat: matches 1. .504 of consensus"

7636..7893

/note="LTR33 repeat: matches 17. .284 of consensus"

9051..9170

/note="MIR repeat: matches 47. .173 of consensus"

9215..9325

/note="L2 repeat: matches 2630. .2750 of consensus"

9362..9413

/note="L2 repeat: matches 2648. .2700 of consensus"

10063..10336

/note="L2 repeat: matches 2453. .2695 of consensus"

10865..10942

/note="L2 repeat: matches 2589. .2670 of consensus"

11068..11234

/note="MIR repeat: matches 66. .239 of consensus"

11287..11381

/note="MIR repeat: matches 151. .249 of consensus"

11816..12075

/note="LTR16B repeat: matches 1. .261 of consensus"

12925..13085

/note="MIR repeat: matches 64. .260 of consensus"

13167..13464

/note="AluX repeat: matches 1. .297 of consensus"

15045..15090

/note="L2 repeat: matches 2696. .2741 of consensus"

15781..15903

/note="MIR repeat: matches 15. .149 of consensus"

16709..16895

/note="MIR repeat: matches 1. .201 of consensus"

17155..17342

/note="L2 repeat: matches 2327. .2496 of consensus"

sequence. The true right end of clone RP4-691N7 is at 100 in this sequence.

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

JOURNAL Submitted (24-FEB-2000) Molecular and Developmental Biology,
Institute of Medical Science, 4-6-1, Shirokanedai, Minato-ku, Tokyo
108-8639, Japan

FEATURES
source
1. .3015
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="18"
/map="18E4"
1. .3015
/gene="Nfatc1"
/note="NFATC"
4. .2157
/gene="Nfatc1"
/note="transcription factor"
/codon_start=1
/product="nuclear factor of activated T cells c"
/protein_id="AA040225.1"
/db_xref="GI:7208618"
/translation="MPNTSPVPKFTPLGPPAAVCGSGETLRPAPPGGTGMKAAEEH
YVSPNVTSTPLTAHSAACHDLOTSTPGISAVPSANHPSPVGGAVDGPSCY
FLSSGNTRNGAPTLPRIETSYLGLHHGSGQFHDVEVDLPCKRSPSTATILH
LPSEIAYRDPSCPLASSLSRSCNSEASYESNYSPASPTSPQSPVSPKTTD
PEGFPSPHGACHLLGSPRSPVSTEDTWLNTQYTSSAIVAAINALTDSLDLGDGP
IKSRKTALEHAPSVALKVEPAGEDLGTTPPTSDPPEEYTFQHLKGAFCOEQYLSVPQ
ASYQWAKPSLSPTSYMSPALDMLPSHSGPYELRVQPKSHRAHYETEGSRG
AVKASAGHPVQLHGYLENEPLTLQIFGTADRLRLPHAFYQVHRITGKTVSTISH
EILNTKFLFELNPNMRAIIDCAGILKRNDSIELRKGTRVRLRVHIPPQNGRTLSQL
VHPQNGRTLSQLVSNPTECSQSAQELPLVEKQSTDPYVIGGKMYLSHNFQDSKVI
DSKVIKVEAPDGHVMEAKTRDLCKPNSLVVEIPFPNRQITSPAQVSFYVCGKRRKRSQYRFTYLP
KRRKRSQYRFTYLPANGNSVFLTSSESELGGFY"

BASE COUNT 705 a 922 c 740 g 648 t
ORIGIN

Query Match 8.3%; Score 22; DB 10; Length 3015;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 243 aggtatgtttgtttgtttgtttg 264
|||||
Db 2221 AGGTATGTTTGTGTTTGTGTTG 2200

RESULT 5
AF049606/c
LOCUS
DEFINITION Mus musculus transcription factor NF-Atc isoform b (Nfatcb) mRNA,
complete cds.
ACCESSION AF049606
VERSION AF049606.1 GI:2952323
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3435)
Sherman,M.A., Powell,D.R., Weiss,D.L. and Brown,M.A.
NF-Atc isoforms are differentially expressed and regulated in
murine T and mast cells
J. Immunol. 165 (5), 2820-2828 (1999)
2 (bases 1 to 3435)
Sherman,M.A., Powell,D.R., Weiss,D.L. and Brown,M.A.
Direct Submission
Submitted (20-FEB-1998) Experimental Pathology, Emory University,
1639 Pierce Drive, Atlanta, GA 30322, USA

FEATURES
source
1. .3435
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="C15"
/cell_type="mast cell"

gene
1. .3435
/gene="Nfatcb"
/note="Nfatcb"
/codon_start=1
/product="transcription factor NF-Atc isoform b"
/protein_id="AAC0505.1"
/db_xref="GI:2952324"
/translation="MTGLEQDPDFDFLEFQSGGAAAAEHSYVSPSVTSTLPL
PTAHSAACHDLOTSTPGISAVPSANHPSPVGGAVDGPSCYFLSSGNTRNGAPT
LESPIETSYLGLHHGSGQFHDVEVDLPCKRSPSTATILHLPSEIAYRDPSCPL
PASSLSRSCNSEASYESNYSPASPTSPQSPVSPKTTDPEGFPSPHGACHLLG
LSPRSPVSTEDTWLNTQYTSSAIVAAINALTDSLDLGDGPVQPKSHRAHYETEG
ALKEPAGEDLGTTPPTSDPPEEYTFQHLKGAFCOEQYLSVPQASYQWAKPSLSPT
SYMSPALDMLPSHSGPYELRVQPKSHRAHYETEGSRGAVKASAGHPVQLHGYLE
NEPLTLQIFGTADRLRLPHAFYQVHRITGKTVSTISH EILNTKFLFELNPNMRA
IIDCAGILKRNDSIELRKGTRVRLRVHIPPQNGRTLSQLVHPQNGRTLSQLVSNP
TECSQSAQELPLVEKQSTDPYVIGGKMYLSHNFQDSKVIKVEAPDGHVMEAKTR
DLCKPNSLVVEIPFPNRQITSPAQVSFYVCGKRRKRSQYRFTYLP KRRKRSQYR
FTYLPANGNSVFLTSSESELGGFY"

BASE COUNT 800 a 990 c 827 g 818 t
ORIGIN

Query Match 8.3%; Score 22; DB 10; Length 3435;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 243 aggtatgtttgtttgtttgtttg 264
|||||
Db 2184 AGGTATGTTTGTGTTTGTGTTG 2163

RESULT 6
AF087434/c
LOCUS
DEFINITION Mus musculus transcription factor NF-Atc isoform a (NF-Atca) mRNA,
complete cds.
ACCESSION AF087434
VERSION AF087434.1 GI:3643194
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3469)
Sherman,M.A., Powell,D.R., Weiss,D.L. and Brown,M.A.
NF-Atc isoforms are differentially expressed and regulated in
murine T and mast cells
J. Immunol. 165 (5), 2820-2828 (1999)
2 (bases 1 to 3469)
Sherman,M.A., Powell,D.R., Weiss,D.L. and Brown,M.A.
Direct Submission
Submitted (26-AUG-1998) Experimental Pathology, Emory University,
1639 Pierce Drive, Atlanta, GA 30322, USA

FEATURES
source
1. .3469
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="CFTL-15"
/cell_type="mast"
1. .3469
/gene="NF-Atca"
/note="NF-Atca"
/codon_start=1
/product="transcription factor NF-Atc isoform a"
/protein_id="AAC36725.1"
/db_xref="GI:3643195"
/translation="MPNTSPVPKFTPLGPPAAVCGSGETLRPAPPGGTGMKAAEEH
YVSPNVTSTPLTAHSAACHDLOTSTPGISAVPSANHPSPVGGAVDGPSCY
FLSSGNTRNGAPTLPRIETSYLGLHHGSGQFHDVEVDLPCKRSPSTATILH


```
/rpt_family="ALU"
complement(51721. .51764)
/rpt_family="I1"
complement(51765. .52058)
/rpt_family="ALU"
complement(52252. .52467)

Query Match      89.4%; Score 237; DB 9; Length 156214;
Best Local Similarity 100.0%; Pred. No. 5.2e-114;
Matches 237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 aaatggtgttcagctgcacatataaagttctactcatccattttttatgacttcta 88
|||||
Db 24902 AAATGGTGTTCAGCTGCCATATAAAGTCTACTCATGCCATTATTTTATGACTTCTA 24961
|||||

QY 89 cgttcagttcacagtagtctgccaattatcgtgggtgaaacttctgtaaatgagatttc 148
|||||
Db 24962 CGTTCAGTTACAAGTATGCTGCAAAATATATCGTGGTGAACACTGTTAAATGAGATTTC 25021
|||||

QY 149 aactgacttagtgatagagttttctcctaagttaattttcacaaatgctatgtttgccaat 208
|||||
Db 25022 AACTGACTTAGTGATAGAGTTTCTTCAAGTTAAATTTTCACAAATGTCATGTTGCAAT 25081
|||||

QY 209 atgaattttctagtcacacattattgtaattaggtatgtttgtttgttttcg 265
|||||
Db 25082 ATGAATTTTCTAGTCAACATATATATGTAATTTAGGTATGTTTGTGTTTGTGC 25138
|||||

RESULT 3
AC099742 172915 bp DNA linear HTG 20-NOV-2001
LOCUS Papio cynocephalus anubis clone RP41-167P22, WORKING DRAFT
DEFINITION AC099742
ACCESSION AC099742.1 GI:17017546
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS olive baboon
SOURCE Papio cynocephalus anubis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopitheciinae; Papio.
REFERENCE 1 (bases 1 to 172915)
AUTHORS Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Graniter, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-Q.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Masiello, C., Mastrian, S.D.,
McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y.,
Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.I., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
Green, E.D.
2 (bases 1 to 172915)
Direct Submission
Submitted (20-NOV-2001) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nhgri.nih.gov
----- Project Information
Center project name: ccy
Center clone name: 167p22
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171178 bases at least Q40
Consensus quality: 171910 bases at least Q30
Consensus quality: 172236 bases at least Q20

Insert size: 130000; agarose-fp
Insert size: 172615; sum-of-contigs
Quality coverage: 12.58x in Q20 bases; agarose-fp
Quality coverage: 9.47x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2438: contig of 2438 bp in length
* 2539: gap of unknown length
* 8133: contig of 5595 bp in length
* 8234: gap of unknown length
* 40378: contig of 32145 bp in length
* 40379: gap of unknown length
* 40479: contig of 132437 bp in length.
FEATURES
Location/Qualifiers
1..172915
/organism="Papio cynocephalus anubis"
/db_xref="taxon:9555"
/clone="RP41-167P22"
/clone_lib="RP41"
1..2438
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
misc_feature 2539..8133
/note="assembly_fragment"
8234..40378
/note="assembly_fragment"
clone_end:T7
vector_side:right"
misc_feature 40479..172915
/note="assembly_fragment"
BASE COUNT 52509 a 31733 c 32277 g 56096 t 300 others
ORIGIN

Query Match      18.1%; Score 48; DB 2; Length 172915;
Best Local Similarity 100.0%; Pred. No. 1.1e-14;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 ttaattttcacaaatgctatgtttgccaatgaaattttcttagtcaa 226
|||||
Db 91762 TTAATTTTCACAAATGTCATGTTGCCATATGAATTTTCTAGTCAA 91809
|||||

RESULT 4
AF239169/c
LOCUS AF239169
DEFINITION Mus musculus nuclear factor of activated T cells c (Nfatc1) mRNA,
complete cds.
ACCESSION AF239169
VERSION AF239169.1 GI:7208617
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3015)
REFERENCE 1 (bases 1 to 3015)
AUTHORS Mori, S., Arai, N. and Arai, K.
TITLE Pan, S., Koyano-Nakagawa, N., Tsuruta, L., Amasaki, Y., Yokota, T.,
Molecular cloning and functional characterization of murine cDNA
encoding transcription factor NFATc
JOURNAL Biochem. Biophys. Res. Commun. 240 (2), 314-323 (1997)
MEDLINE 98049829
REFERENCE 2 (bases 1 to 3015)
AUTHORS Pan, S.
TITLE Direct Submission
```

VECTOR: pBELO

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of H_RG016J04;
 actual end is at 136214 of H_RG016J04. The orientation of this
 clone is unknown.

This clone contains STS SWSS2784 (NID:gl113580) and SWSS893
 (NID:g454733).

FEATURES

source Location/Qualifiers

1. 136214
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7q21"
 /clone="RG016J04"

/clone_lib="CITB-978SK-B"
 15. 40
 /rpt_family="L1"

repeat_region

326. 621

repeat_region

/rpt_family="ALU"

repeat_region

complement(977. 1499)

repeat_region

/rpt_family="L1"

repeat_region

/rpt_family="L1"

repeat_region

/rpt_family="L1"

repeat_region

complement(8071. 8347)

repeat_region

/rpt_family="ALU"

repeat_region

complement(9406. 9975)

repeat_region

/rpt_family="L1"

repeat_region

complement(10000. 11285)

repeat_region

/rpt_family="L1"

repeat_region

complement(11315. 11984)

repeat_region

/rpt_family="L1"

repeat_region

/rpt_family="L1"

repeat_region

complement(12301. 13893)

repeat_region

/rpt_family="L1"

repeat_region

/rpt_family="L1"

repeat_region

complement(14777. 14838)

repeat_region

/rpt_family="L1"

repeat_region

15715. 15767

misc_feature

/note="similar to human EST T02878 (NID:g3141119)"

misc_feature

19436. 19497

misc_feature

/note="similar to human EST AAL23941 (NID:g1682616)

misc_feature

mq22f09.r1

repeat_region

complement(25113. 25148)

repeat_region

/rpt_family="L1"

repeat_region

complement(25561. 25578)

repeat_region

/rpt_family="L1"

repeat_region

complement(26832. 27124)

repeat_region

/rpt_family="ALU"

misc_feature

27125. 27383

misc_feature

/note="similar to human EST AA151807 (NID:gl720502)

misc_feature

2139g02.r1

misc_feature

27125. 27544

misc_feature

/note="similar to human EST AA151796 (NID:gl720491)

misc_feature

2139c02.r1

misc_feature

complement(27178. 27600)

misc_feature

/note="similar to human EST AA149579 (NID:gl720380)

misc_feature

2139c02.s1

misc_feature

complement(27441. 27890)

misc_feature

/note="similar to human EST N52554 (NID:gl193720)

misc_feature

yv36a11.s1

misc_feature

27686. 27856

misc_feature

/note="similar to human EST W32120 (NID:gl1313113)

misc_feature

2b97c08.r1

misc_feature

27686. 27838

misc_feature

/note="similar to human EST W32085 (NID:gl313105)
 zb97b08.r1"

misc_feature

complement(27733. 28143)

misc_feature

/note="similar to human EST N59831 (NID:gl203721)
 yz77a10.s1"

misc_feature

complement(27838. 28143)

/note="similar to human EST W31561 (NID:gl312680)
 zb97b08.s1"

misc_feature

complement(27887. 28143)

/note="similar to human EST N59830 (NID:gl203720)
 yz77a09.s1"

misc_feature

complement(27997. 28144)

/note="similar to human EST W31628 (NID:gl312688)
 zb97c08.s1"

misc_feature

complement(28972)

/note="similar to human EST R09339 (NID:g761262)
 yf26d05.r1"

misc_feature

complement(29587. 29853)

/note="similar to human EST R09227 (NID:g761150)
 yf26d05.s1"

misc_feature

complement(30146. 30314)

/note="similar to human EST N78042 (NID:gl240743)
 yv71e06.r1"

misc_feature

complement(31170. 31350)

/note="similar to human EST N58451 (NID:gl202341)
 yv71e06.s1"

repeat_region

31970. 32006

repeat_region

/rpt_family="L1"

repeat_region

32263. 32563

repeat_region

/rpt_family="ALU"

repeat_region

complement(34737. 34778)

repeat_region

36558. 36709

repeat_region

/rpt_family="ALU"

repeat_region

complement(37630. 37740)

repeat_region

/rpt_family="PTR"

repeat_region

39839. 40125

repeat_region

/rpt_family="ALU"

repeat_region

40126. 40157

repeat_region

/rpt_family="L1"

repeat_region

complement(41340. 41706)

repeat_region

/rpt_family="ALU"

repeat_region

complement(41904. 41933)

repeat_region

/rpt_family="L1"

repeat_region

42348. 42409

repeat_region

/rpt_family="ALU"

repeat_region

complement(42942. 42982)

repeat_region

/rpt_family="L1"

repeat_region

complement(44380. 44423)

repeat_region

/rpt_family="L1"

repeat_region

44659. 44954

repeat_region

/rpt_family="ALU"

repeat_region

45422. 45448

repeat_region

/rpt_family="L1"

repeat_region

complement(45615. 45669)

repeat_region

/rpt_family="L1"

repeat_region

complement(45672. 45963)

repeat_region

/rpt_family="ALU"

repeat_region

complement(45964. 46597)

repeat_region

/rpt_family="L1"

repeat_region

46100. 46516

repeat_region

/rpt_family="L1"

repeat_region

48370. 48785

repeat_region

/rpt_family="L1"

repeat_region

complement(49577. 50216)

repeat_region

/rpt_family="L1"

repeat_region

complement(50540. 50791)

repeat_region

/rpt_family="THE"

repeat_region

complement(50820. 50985)

repeat_region

/rpt_family="THE"

repeat_region

complement(51124. 51156)

repeat_region

/rpt_family="L1"

repeat_region

51405. 51691

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:57:55 ; Search time 4356.39 Seconds
(without alignments)
1272.965 Million cell updates/sec

Title: US-09-802-520-9
Perfect score: 265
Sequence: 1 ccgagagggtcacagtaat.....tatgtttgttttttttgc 265

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_em.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_on.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description

c	1	262	98.9	162928	2	AC104475	
c	2	237	89.4	156214	9	HSAC002064	
c	3	48	18.1	172915	2	AC099742	
c	4	22	8.3	3015	10	AF239169	
c	5	22	8.3	3435	10	AF049606	
c	6	22	8.3	3469	10	AF087434	
c	7	22	8.3	64855	9	AL139219	
c	8	22	8.3	171483	2	AC009278	
c	9	22	8.3	179408	2	AC027234	
c	10	21	7.9	8253	9	AF198614	
c	11	21	7.9	10917	9	AL592167	
c	12	21	7.9	109395	2	AC013742	
c	13	21	7.9	110000	2	AL390202_06	
c	14	21	7.9	135005	9	HS860F19	
c	15	21	7.9	140554	2	AL645757	
c	16	21	7.9	157636	9	AL360011	
c	17	21	7.9	158978	2	AC095026	
c	18	21	7.9	159287	2	AC108713	
c	19	21	7.9	171681	2	AC091541	
c	20	21	7.9	174327	2	AC068159	
c	21	21	7.9	176550	9	AL356356	
c	22	21	7.9	184455	9	AC010740	
c	23	21	7.9	184855	2	AL606744	
c	24	21	7.9	185309	2	AC108687	
c	25	21	7.9	186044	2	AC084099	
c	26	21	7.9	207815	2	AC053497	
c	27	21	7.9	214651	2	AC091464	
c	28	21	7.9	235564	2	AC022062	
c	29	21	7.9	239480	2	AC079818	
c	30	20	7.5	1433	8	AF170966	
c	31	20	7.5	2876	8	PNCBT	
c	32	20	7.5	7189	6	AX344619	
c	33	20	7.5	17727	3	U61945	
c	34	20	7.5	26727	2	AC107535	
c	35	20	7.5	30407	2	AC100899	
c	36	20	7.5	35705	9	AC010522	
c	37	20	7.5	38342	6	AX251504	
c	38	20	7.5	38342	6	AX344503	
c	39	20	7.5	39880	9	AC010645	
c	40	20	7.5	44764	2	AC100016	
c	41	20	7.5	48731	9	AC004211	
c	42	20	7.5	52614	10	AC006056	
c	43	20	7.5	61054	9	AL133323	
c	44	20	7.5	61885	2	AL102640	
c	45	20	7.5	63662	2	AC025398	

ALIGNMENTS

RESULT 1
AC104475/c
LOCUS
DEFINITION
AC104475 162928 bp DNA linear HTG 12-DEC-2001
Pan troglodytes clone RP43-120K11, WORKING DRAFT SEQUENCE, 6
unordered pieces.
AC104475
AC104475.1 GI:17530717
HTG: HTGS_PHASE1; HTGS_DRAFT.
chimpanzee.
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
1 (bases 1 to 162928)
AUTHORS
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Mastello,C., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Stantripop,S.,
Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsugeon,C., Vogt,J.L.,
Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and
Green,E.D.
TITLE
NISC Comparative Sequencing Initiative

```

/clone="IMAGE:4869439"
/clone_lib="NIH_MGC_49"
/tissue_type="melanotic melanoma, high MDR (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: POB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library. I"
BASE COUNT      811 a 290 c 298 g 18 t
ORIGIN

Query Match      9.3%; Score 19; DB 10; Length 1417;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 gttcttgctgctgttttc 44
      |||||||
Db 459 gttcttgctgctgttttc 441

RESULT 14
LOCUS      BI467930      143 bp      mRNA      linear      EST 22-AUG-2001
DEFINITION 389764 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION  BI467930
VERSION    BI467930.1 GI:15280808
KEYWORDS   EST.
SOURCE     pig.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 143)
AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.,
            and Keele,J.W.
TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
JOURNAL   Unpublished (2000)
COMMENT   Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAACAGCTATGACCAT
            BACKWARD: GTTTCCTCCAGTCACGACG
            Plate: 143 row: I column: 19
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES   Location/Qualifiers
            source
            1..143
            /organism="Sus scrofa"
            /db_xref="taxon:9823"
            /clone_lib="MARC 2P1G"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
            Library made from pooled tissue from testis, ovary,
            endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      37 a 44 c 31 g 31 t
ORIGIN

Query Match      8.8%; Score 18; DB 10; Length 143;

```

```

Best Local Similarity 100.0%; Pred. No. 43;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 tgtcttgctgctgtttt 42
      |||||||
Db 130 tgtcttgctgctgtttt 113

RESULT 15
LOCUS      AW308401/c    218 bp      mRNA      linear      EST 08-JAN-2001
DEFINITION 3167 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION  AW308401
VERSION    AW308401.1 GI:6720764
KEYWORDS   EST.
SOURCE     pig.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 218)
AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.,
            and Keele,J.W.
TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
JOURNAL   Unpublished (2000)
COMMENT   Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAACAGCTATGACCAT
            BACKWARD: GTTTCCTCCAGTCACGACG
            Plate: 138 row: I column: 16
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES   Location/Qualifiers
            source
            1..218
            /organism="Sus scrofa"
            /db_xref="taxon:9823"
            /clone_lib="MARC 1P1G"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
            Library made from pooled tissue from day 11, 13, 15, 20,
            and 30 embryos."
BASE COUNT      53 a 70 c 57 g 38 t
ORIGIN

Query Match      8.8%; Score 18; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 tgtcttgctgctgtttt 42
      |||||||
Db 216 tgtcttgctgctgtttt 199

Search completed: September 20, 2002, 08:33:58
Job time: 8722 sec

```


/cell_type="Sperm"	
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2: HindIII"	
BASE COUNT	148 a 83 c 95 g 140 t
ORIGIN	
Query Match 9.3%; Score 19; DB 12; Length 466;	
Best Local Similarity 100.0%; Pred. No. 14;	
Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	52 tgtaattctgggaagatt 70
Db	213 TGTAATTCTGGGAAGATT 231
RESULT 10	
AQ820554	
LOCUS	551 bp DNA linear GSS 26-AUG-1999
DEFINITION	HS 5443_A2_F08_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1019 Col=16 Row=K, DNA sequence.
ACCESSION	AQ820554
VERSION	AQ820554.1 GI:5782947
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 551)
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@delong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 1019 row: K column: 16 Seq primer: SP6 Class: BAC ends High quality sequence stop: 551. Location/Qualifiers 1..551 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="Plate=1019 Col=16 Row=K" /clone_lib="RPCI-11 Human Male BAC Library" /sex="male" /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites" BASE COUNT 174 a 128 c 117 g 125 t ORIGIN 7 others
Query Match 9.3%; Score 19; DB 12; Length 551;	
Best Local Similarity 100.0%; Pred. No. 15;	
Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ACCESSION	AI208199
VERSION	AI208199.1
KEYWORDS	GI:3770141
SOURCE	EST.
ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 423) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Insert Length: 380 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 224. Location/Qualifiers 1..423 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1835094" /clone_lib="Soares_testis_NHT" /sex="male" /lab_host="DH10B" /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I::oligo(dT) primer [5', TGTTACCAATCTGAAGTGAGCGGCCCCCAATTGTTCCTTTT 3'] Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT	88 a 96 c 62 g 177 t
ORIGIN	
Query Match	9.3%; Score 19; DB 9; Length 423;
Best Local Similarity	100.0%; Pred.No.14;
Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	101 taaacgaatgaagaag 119
Db	36 TAAACGAATTAAGAAG 18
RESULT	8
LOCUS	AQ063518/c
DEFINITION	HS_2185_A2_C05_MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2185 Col=10 Row=E, DNA sequence.
ACCESSION	AQ063518
VERSION	AQ063518.1
KEYWORDS	GI:3378776
SOURCE	GSS. human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 444) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L. Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
REFERENCE	
AUTHORS	
TITLE	

BASE COUNT	688 a 528 c 517 g 740 t
ORIGIN	
Query Match	9.8%; Score 20; DB 11; Length 2473;
Best Local Similarity	100.0%; Pred.No.5;
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	94 cgaaagctgtaaacgaattaa 113
Db	681 CGAAGCTAAACGAATTAA 700
RESULT	6
LOCUS	AG024836
DEFINITION	Oryza sativa DNA, 3' flanking sequence of Tos17 insertion, clone:T2942T, genomic survey sequence.
ACCESSION	AG024836
VERSION	AG024836.1
KEYWORDS	GI:7683500
SOURCE	GSS; GSS (genome survey sequence). Oryza sativa (sub.species:japonica, strain:NC0384, cultivar:Nipponbare) DNA, clone_lib:PCR product directly amplified from rice genomic DNA clone:T2942T.
ORGANISM	Oryza sativa Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza. 1 (sites) Miyao,A. and Hirochika,H. Rice insertion mutants Unpublished (1999) 2 (bases 1 to 289) Miyao,A., Miyazaki,A., Yamashita,Y. and Hirochika,H. Direct Submission Submitted (25-Oct-1999) to the DDBJ/EMBL/GenBank databases. Akio Miyao, National Institute of Agrobiological Resources, Molecular Genetics; 2-1-2, Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:milyao@affrc.go.jp, URL: http://www.abr.affrc.go.jp/ , Tel:81-298-38-7006, Fax:81-298-38-7006)
FEATURES	Location/Qualifiers 1..289 /organism="Oryza sativa" /cultivar="Nipponbare" /strain="NC0384" /sub_species="japonica" /db_xref="taxon:4530" /clone="T2942T" /clone_lib="PCR product directly amplified from rice genomic DNA" /note="Sequence group name: T2942T.-The 3' end of retrotransposon Tos17 was found immediately upstream of this sequence."
BASE COUNT	92 a 48 c 67 g 82 t
ORIGIN	
Query Match	9.3%; Score 19; DB 12; Length 289;
Best Local Similarity	100.0%; Pred.No.13;
Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	113 agaaagctggggaag 131
Db	26 AGAAGCTGGGAAGAG 44
RESULT	7
LOCUS	AI208199/c
DEFINITION	qg53ci2.xl1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1839094 similar to SW:RA54_HUMAN P46100 X-LINKED HELICASE II ;, mrna sequence.

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LHAM9851 row: d column: 08
High quality sequence stop: 660.
FEATURES
source
1. .948
Location/Qualifiers
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4238935"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library. I"

BASE COUNT 228 a 244 c 231 g 245 t
ORIGIN

Query Match 9.8%; Score 20; DB 10; Length 948;
Best Local Similarity 100.0%; Pred. No. 4.; 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0

QY 94 cgaagctaaaacgaattaa 113
|||||
Db 356 CGAAGCTAAACGAATTAA 375
|||||

RESULT 5
AK015015
LOCUS
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4921538B17:homolog to CDNA FLJ14101 FIS, CLONE MAMMA1000859, full insert sequence.
ACCESSION AK015015
VERSION 1 GI:12853186
KEYWORDS HTG; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:4921538B17.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (sites)
Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE 2 (sites)
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL 20499374
MEDLINE 11042159
PUBMED 11042159

REFERENCE 3 (sites)
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yanamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED 11076861

4 (sites)
THE RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
5 (bases 1 to 2473)
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koyama,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGATTCGATTAATTAATCCGCCGCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised longer than 7 kb was selected before cloning. Vector: a modified pluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
Location/Qualifiers
1. .2473
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGD:MGI:1907109"
/db_xref="taxon:10090"
/clone="4921538B17"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
390. .788
/note="data source:SPTR, source key:Q9H7Y1, evidence:ISS homolog to CDNA FLJ14101 FIS, CLONE MAMMA1000859 putative"
/codon_start=1
/protein_id="BAB29672.1"
/db_xref="GI:12853187"
/translation="MYISFGIMSLGLSLAVTSPISVSNALNWRFSFIQTSLGYVA LLITFFVLIVGWKRAFAEYRYTPNFVLAIVLPSIVILGKMLILLPISRLKLR IKKGWKSQFLDECMGCAVPHLSERVTVM"
2444. .2449
polyA_signal
note="putative"
2473
polyA_site

```

Query Match      14.7%; Score 30; DB 10; Length 558;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ttatataccaccacaaacttctgttctgc 35
    |||
Db 143 TTTATACACCAACAACTTGTCTGTCTC 172
    |||

RESULT 2
A0236699      579 bp DNA linear GSS 21-APR-1999
DEFINITION    RPC111-68C15.TK RPC1-11 Homo sapiens genomic clone RPC1-11-68C15,
DNA sequence.
ACCESSION    A0236699
VERSION      A0236699.1 GI:3668990
KEYWORDS     GSS.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 579)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Other GSSs: RPC111-68C15.TJ
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (http://resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1..579
/organism="Homo sapiens"
/db_xref="GDB:7525790"
/db_xref="taxon:9606"
/clone="RPC1-11-68C15"
/clone_lib="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

BASE COUNT    170 a 114 c 144 g 151 t
ORIGIN

Query Match      10.3%; Score 21; DB 12; Length 579;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 ttgtcttgcctgtgttgg 43
    |||
Db 49 TTTGTCTGCTGCTGCTGTTG 69
    |||

RESULT 3
A0836131/c
LOCUS
DEFINITION    HS_5524_A2_H09_T7A RPC1-11 Human Male BAC library Homo sapiens
genomic clone Plate=1100 Col=18 Row=0, DNA sequence.
A0836131
LOCUS
DEFINITION    A0836131.1 GI:5806005
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 526)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (http://resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1100 row: 0 column: 18
Seq primer: T7
Class: BAC ends
High quality sequence stop: 526.
Location/Qualifiers
1..526
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1100 Col=18 Row=0"
/clone_lib="RPC1-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"

BASE COUNT    142 a 121 c 88 g 170 t 5 others
ORIGIN

Query Match      9.8%; Score 20; DB 12; Length 526;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 112 aagaaagcctgggaaag 131
    |||
Db 371 AGAAGAGCTGGGGAAGAG 352
    |||

RESULT 4
BF784438
LOCUS
DEFINITION    602110826f1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4238935
5', mRNA sequence.
ACCESSION    BF784438
VERSION      BF784438.1 GI:12089474
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 948)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 08:33:54 ; Search time 5053.42 seconds
(without alignments)
544.855 Million cell updates/sec

Title: US-09-802-520-8

Perfect score: 204

Sequence: 1 cagagtttatacaccaccaa.....gggtcacagtaatgggatga 204

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	14.7	558	10	BM431438
2	21	10.3	579	12	AQ236699
3	20	9.8	526	12	AQ836131
4	20	9.8	948	10	BF784438
5	20	9.8	2473	11	AK015015
6	19	9.3	289	12	AG024836
7	19	9.3	423	9	A1208199
8	19	9.3	444	12	AQ063518
9	19	9.3	466	12	AQ056717
10	19	9.3	551	12	AQ820554
11	19	9.3	629	12	AQ345537
12	19	9.3	1065	10	BF337390
13	19	9.3	1417	10	BG765708
14	18	8.8	143	10	BI467930
15	18	8.8	218	9	AW308401
16	18	8.8	288	9	AW480075
17	18	8.8	355	10	BI118284

18	18	8.8	543	12	BH110402
19	18	8.8	546	9	BE031241
20	18	8.8	633	9	AI054511
21	18	8.8	888	12	AQ739010
22	17	8.3	138	10	R79044
23	17	8.3	198	9	AV745201
24	17	8.3	207	10	BG313609
25	17	8.3	238	10	F13874
26	17	8.3	245	10	BE930650
27	17	8.3	248	12	BH609699
28	17	8.3	259	10	BF746375
29	17	8.3	261	10	D78884
30	17	8.3	267	10	N40676
31	17	8.3	270	10	BE930663
32	17	8.3	276	10	BG000486
33	17	8.3	291	10	D58789
34	17	8.3	294	10	T48916
35	17	8.3	297	9	BB495914
36	17	8.3	299	12	AQ478263
37	17	8.3	303	10	BG013160
38	17	8.3	315	9	AV746344
39	17	8.3	321	9	BB320148
40	17	8.3	330	10	CI8403
41	17	8.3	336	10	D79076
42	17	8.3	336	10	R79875
43	17	8.3	340	9	BE155764
44	17	8.3	347	10	R34458
45	17	8.3	348	10	R33236

ALIGNMENTS

RESULT 1

BM431438	BM431438	558 bp	linear	EST 31-JAN-2002
LOCUS	1DU016F03	Bos taurus Duodenum #1 library	Bos taurus	cdna, mRNA
DEFINITION	sequence.			
ACCESSION	BM431438			
VERSION	BM431438.1	GI:18453160		
KEYWORDS	EST.			
SOURCE	COW.			
ORGANISM	Bos taurus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.			
AUTHORS	1 (bases 1 to 558) Hansen, C., Fu, A., Meng, Y., Li, C., Okine, E., Senses, C.W., Gordon, P.M.K. and Moore, S.S.			
TITLE	Gene Expression Profiling of the Bovine Gastrointestinal Tract			
JOURNAL	Unpublished (2002)			
COMMENT	Contact: Dr. Stephen Moore Beef Genomics Laboratory Dept of AFNS, University of Alberta 410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada Tel: 780 492 0169 Fax: 780 492 4265 Email: smoores@afns.ualberta.ca Insert Length: 558 Std Error: 0.00 POLYA-No.			
FEATURES	Location/Qualifiers			
source	1..558 /organism="Bos taurus" /db_xref="taxon:9913" /clone_lib="Bos taurus Duodenum #1 library" /tissue_type="Smooth muscle" /cell_type="Simple columnar epithelial" /dev_stage="Young adult" /lab_host="X11-BlueMRF/strain" /note="Organ: Intestine/duodenum; Vector: Uni-22APXR; Site_1: ECORI; Site_2: Xho I"			
BASE COUNT	152 a	109 c	108 g	189 t
ORIGIN				

THIS PAGE BLANK (USPTO)

APPLICANT: MOYLE, William R.
TITLE OF INVENTION: SINGLE-CHAIN FORMS OF THE
TITLE OF INVENTION: GLYCOPROTEIN HORMONE QUARTET
NUMBER OF SEQUENCES: 83
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Avenue, NW, suite 5500
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/282,357
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/918,288
FILING DATE: 25 AUG-1997
APPLICATION NUMBER: 08/853,524
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: 08/199,382
FILING DATE: 18-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 29500-20050.25
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-887-1500
TELEFAX: 202-887-0763
TELEX:
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 693 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-282-357-28

Query Match 8.3% Score 17; DB 4; Length 693;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 184 aggtcacagtaatggg 200
|||||
Db 87 AGGTCACAGTAATGGG 71

Search completed: September 20, 2002, 09:51:18
Job time: 11866 sec

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-887-1500
 TELEFAX: 202-887-0763
 TELEX:
 INFORMATION FOR SEQ ID NO: 31:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 681 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-918-288-31

Query Match 8.3%; Score 17; DB 4; Length 681;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 agggtcacagtaatggg 200
 Db 87 AGGGTCACAGTAATGGG 71

RESULT 13
 US-09-282-357-31/c
 ; Sequence 31, Application US/09282357
 ; Patent No. 6242580
 ; GENERAL INFORMATION:
 ; APPLICANT: BOIME, Irving
 ; APPLICANT: MOYLE, William R.
 ; TITLE OF INVENTION: SINGLE-CHAIN FORMS OF THE
 ; TITLE OF INVENTION: GLYCOPROTEIN HORMONE QUARTET
 ; NUMBER OF SEQUENCES: 83
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORRISON & FOERSTER
 ; STREET: 2000 Pennsylvania Avenue, NW, suite 5500
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20006-1888
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/282,357
 ; FILING DATE:
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/918,288
 ; FILING DATE: 25 AUG-1997
 ; APPLICATION NUMBER: 08/853,524
 ; FILING DATE: 09-MAY-1997
 ; APPLICATION NUMBER: 08/199,382
 ; FILING DATE: 18-FEB-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Murashige, Kate H
 ; REGISTRATION NUMBER: 29,959
 ; REFERENCE/DOCKET NUMBER: 29500-20050.25
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-887-1500
 ; TELEFAX: 202-887-0763
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 31:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 681 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; US-09-282-357-31

Query Match 8.3%; Score 17; DB 4; Length 681;

Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 184 agggtcacagtaatggg 200
 Db 87 AGGGTCACAGTAATGGG 71

RESULT 14
 US-08-918-288-28/c
 ; Sequence 28, Application US/08918288
 ; Patent No. 6238890
 ; GENERAL INFORMATION:
 ; APPLICANT: BOIME, Irving
 ; APPLICANT: MOYLE, William R.
 ; TITLE OF INVENTION: SINGLE-CHAIN FORMS OF THE
 ; TITLE OF INVENTION: GLYCOPROTEIN HORMONE QUARTET
 ; NUMBER OF SEQUENCES: 83
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORRISON & FOERSTER
 ; STREET: 2000 Pennsylvania Avenue, NW, suite 5500
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20006-1888
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/918,288
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/282,357
 ; FILING DATE:
 ; APPLICATION NUMBER: 08/853,524
 ; FILING DATE: 09-MAY-1997
 ; APPLICATION NUMBER: 08/199,382
 ; FILING DATE: 18-FEB-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Murashige, Kate H
 ; REGISTRATION NUMBER: 29,959
 ; REFERENCE/DOCKET NUMBER: 29500-20050.25
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-887-1500
 ; TELEFAX: 202-887-0763
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 28:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 693 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; US-08-918-288-28

Query Match 8.3%; Score 17; DB 4; Length 693;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 agggtcacagtaatggg 200
 Db 87 AGGGTCACAGTAATGGG 71

RESULT 15
 US-09-282-357-28/c
 ; Sequence 28, Application US/09282357
 ; Patent No. 6242580
 ; GENERAL INFORMATION:
 ; APPLICANT: BOIME, Irving

TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..348
US-09-322-676-4

Query Match 8.3% Score 17; DB 3; Length 351;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 184 agggtcacagtaatggg 200
|||||
Db 271 AGGGTCACAGTAATGGG 287

RESULT 10

US-09-466-036A-4
; Sequence 4, Application US/09466036A
; Patent No. 6281197
; GENERAL INFORMATION:

APPLICANT: Floriewicz, Robert Z.
TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:

ADDRESSEE: SED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/466.036A

FILING DATE: 17-Dec-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/211,290

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 200124.401D1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 351 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: 1..348

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-466-036A-4

Query Match 8.3% Score 17; DB 4; Length 351;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 184 agggtcacagtaatggg 200
|||||
Db 271 AGGGTCACAGTAATGGG 287

RESULT 11

US-09-451-905-4

; Sequence 4, Application US/09451905

; Patent No. 6306613

; GENERAL INFORMATION:

APPLICANT: Robert Z. Floriewicz

APPLICANT: Andrew Baird

APPLICANT: Dale E. Warnock

TITLE OF INVENTION: MODULATORS OF LEADERLESS PROTEIN EXPORT
AND METHODS FOR IDENTIFYING AND USING THE SAME

FILE REFERENCE: 200124.402C4

CURRENT APPLICATION NUMBER: US/09/451,905

CURRENT FILING DATE: 1999-12-01

NUMBER OF SEQ ID NOS: 48

SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 4

LENGTH: 351

TYPE: DNA

ORGANISM: Homo sapien

US-09-451-905-4

Query Match 8.3% Score 17; DB 4; Length 351;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 184 agggtcacagtaatggg 200
|||||
Db 271 agggtcacagtaatggg 287

RESULT 12

US-08-918-288-31/c

; Sequence 31, Application US/08918288

; Patent No. 6238890

; GENERAL INFORMATION:

APPLICANT: BOIME, Irving

APPLICANT: MOYLE, William R.

TITLE OF INVENTION: SINGLE-CHAIN FORMS OF THE

TITLE OF INVENTION: GLYCOPROTEIN HORMONE QUARTET

NUMBER OF SEQUENCES: 83

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 2000 Pennsylvania Avenue, NW, suite 5500

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20006-1888

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/918,288

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/282,357

FILING DATE:

APPLICATION NUMBER: 08/853,524

FILING DATE: 09-MAY-1997

APPLICATION NUMBER: 08/199,382

FILING DATE: 18-FEB-1994

ATTORNEY/AGENT INFORMATION:

NAME: Murashige, Kate H

REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 29500-20050.25


```
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 agggtcacagtaatggg 200
Db 271 AGGGTCACAGTAATGGG 287

RESULT 7
US-09-211-290-4
; Sequence 4, Application US/09211290
; Patent No. 6071885
; GENERAL INFORMATION:
; APPLICANT: FlorKiewicz, Robert Z.
; TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/211,290
; FILING DATE: 12-DEC-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 200124.401D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..348
; US-09-211-290-4

Query Match 8.3%; Score 17; DB 3; Length 351;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 agggtcacagtaatggg 200
Db 271 AGGGTCACAGTAATGGG 287

RESULT 8
US-09-030-613-4
; Sequence 4, Application US/09030613
; Patent No. 6083706
; GENERAL INFORMATION:
; APPLICANT: FlorKiewicz, Robert Z.
; APPLICANT: Baird, J. Andrew
; TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle

Query Match 8.3%; Score 17; DB 3; Length 351;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 agggtcacagtaatggg 200
Db 271 AGGGTCACAGTAATGGG 287

RESULT 9
US-09-322-676-4
; Sequence 4, Application US/09322676
; Patent No. 6107283
; GENERAL INFORMATION:
; APPLICANT: FlorKiewicz, Robert Z.
; TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/322,676
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/211,290
; FILING DATE: 12-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 200124.401D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
```

ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/918,288
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/282,357
FILING DATE:
APPLICATION NUMBER: 08/853,524
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: 08/199,382
FILING DATE: 18-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 29500-20050.25
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-887-1500
TELEFAX: 202-887-0763
TELEX:
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-918-288-34

Query Match 8.3%; Score 17; DB 4; Length 317;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 agggtcacagtaatggg 200
|||||
DB 92 AGGTCACAGTAATGGG 76

RESULT 5
US-09-282-357-34/c
Sequence 34, Application US/09282357
Patent No. 6242580
GENERAL INFORMATION:
APPLICANT: BOIME, Irving
APPLICANT: MOYLE, William R.
TITLE OF INVENTION: SINGLE-CHAIN FORMS OF THE
TITLE OF INVENTION: GLYCOPROTEIN HORMONE QUARTET
NUMBER OF SEQUENCES: 83
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Avenue, NW, suite 5500
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/282,357
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/918,288
FILING DATE: 25 AUG-1997
APPLICATION NUMBER: 08/853,524

FILING DATE: 09-MAY-1997
APPLICATION NUMBER: 08/199,382
FILING DATE: 18-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 29500-20050.25
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-887-1500
TELEFAX: 202-887-0763
TELEX:

INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-282-357-34

Query Match 8.3%; Score 17; DB 4; Length 317;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 agggtcacagtaatggg 200
|||||
DB 92 AGGTCACAGTAATGGG 76

RESULT 6
US-08-599-895-4
Sequence 4, Application US/08599895
Patent No. 5891855
GENERAL INFORMATION:
APPLICANT: Florjanczyk, Robert Z.
TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/599,895
FILING DATE: 31-JAN-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: No. 5891855tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 760100.416
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 351 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..348
US-08-599-895-4

Query Match 8.3%; Score 17; DB 2; Length 351;

Db 420 TTTATACACCAAACTTGTCTTGTCTC 449
|||||

RESULT 2

US-08-918-288-32
; Sequence 32, Application US/08918288
; Patent No. 6238890
; GENERAL INFORMATION:
; APPLICANT: BOIME, Irving
; APPLICANT: MOYLE, William R.
; TITLE OF INVENTION: SINGLE-CHAIN FORMS OF THE
; TITLE OF INVENTION: GLYCOPROTEIN HORMONE QUARTET
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW, suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/918,288
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/282,357
; FILING DATE:
; APPLICATION NUMBER: 08/853,524
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: 08/199,382
; FILING DATE: 18-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 29500-20050.25
; TELEPHONE: 202-887-1500
; TELEFAX: 202-887-0763
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 312 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..303
; OTHER INFORMATION:
; US-08-918-288-32

Query Match 8.3%; Score 17; DB 4; Length 312;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 agggtcacagtaatggg 200
|||||

Db 226 AGGTCACAGTAATGGG 242

RESULT 3

US-09-282-357-32
; Sequence 32, Application US/09282357
; Patent No. 6242580
; GENERAL INFORMATION:
; APPLICANT: BOIME, Irving

; APPLICANT: MOYLE, William R.
; TITLE OF INVENTION: SINGLE-CHAIN FORMS OF THE
; TITLE OF INVENTION: GLYCOPROTEIN HORMONE QUARTET
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW, suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/282,357
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/918,288
; FILING DATE: 25 AUG-1997
; APPLICATION NUMBER: 08/853,524
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: 08/199,382
; FILING DATE: 18-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 29500-20050.25
; TELEPHONE: 202-887-1500
; TELEFAX: 202-887-0763
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 312 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..303
; OTHER INFORMATION:
; US-09-282-357-32

Query Match 8.3%; Score 17; DB 4; Length 312;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 agggtcacagtaatggg 200
|||||

Db 226 AGGTCACAGTAATGGG 242

RESULT 4

US-08-918-288-34/c
; Sequence 34, Application US/08918288
; Patent No. 6238890
; GENERAL INFORMATION:
; APPLICANT: BOIME, Irving
; APPLICANT: MOYLE, William R.
; TITLE OF INVENTION: SINGLE-CHAIN FORMS OF THE
; TITLE OF INVENTION: GLYCOPROTEIN HORMONE QUARTET
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW, suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:51:17 ; Search time 139.75 Seconds
(without alignments)
358.563 Million cell updates/sec

Title: US-09-802-520-8
Perfect score: 204
Sequence: 1 cagagtttatcacaccacaa.....gggtcacagtaatgggatga 204

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	14.7	1213	3	US-09-083-521-3
2	17	8.3	312	4	US-08-918-288-32
3	17	8.3	312	4	US-09-282-357-32
C 4	17	8.3	317	4	US-08-918-288-34
C 5	17	8.3	317	4	US-09-282-357-34
6	17	8.3	351	2	US-08-599-895-4
7	17	8.3	351	3	US-09-211-290-4
8	17	8.3	351	3	US-09-030-613-4
9	17	8.3	351	3	US-09-322-676-4
10	17	8.3	351	4	US-09-466-036A-4
11	17	8.3	351	4	US-09-451-905-4
C 12	17	8.3	681	4	US-08-918-288-31
C 13	17	8.3	681	4	US-09-282-357-31
C 14	17	8.3	693	4	US-08-918-288-28
C 15	17	8.3	693	4	US-09-282-357-28
C 16	17	8.3	702	4	US-08-918-288-13
C 17	17	8.3	702	4	US-09-282-357-13
C 18	17	8.3	707	4	US-08-918-288-29
19	17	8.3	707	4	US-09-282-357-29
C 20	17	8.3	717	4	US-08-918-288-7
C 21	17	8.3	717	4	US-08-918-288-22
C 22	17	8.3	717	4	US-08-918-288-25
C 23	17	8.3	717	4	US-09-282-357-7
C 24	17	8.3	717	4	US-09-282-357-22
C 25	17	8.3	717	4	US-09-282-357-25
C 26	17	8.3	718	4	US-08-918-288-10
C 27	17	8.3	718	4	US-09-282-357-10

28	17	8.3	719	4	US-08-918-288-26	Sequence 26, Appl
29	17	8.3	719	4	US-09-282-357-26	Sequence 26, Appl
C 30	17	8.3	726	4	US-08-918-288-16	Sequence 16, Appl
C 31	17	8.3	726	4	US-08-918-288-19	Sequence 19, Appl
C 32	17	8.3	726	4	US-09-282-357-16	Sequence 16, Appl
C 33	17	8.3	726	4	US-09-282-357-19	Sequence 19, Appl
34	17	8.3	728	4	US-08-918-288-11	Sequence 11, Appl
35	17	8.3	728	4	US-09-282-357-11	Sequence 11, Appl
36	17	8.3	743	4	US-08-918-288-5	Sequence 5, Appl
37	17	8.3	743	4	US-08-918-288-20	Sequence 20, Appl
38	17	8.3	743	4	US-08-918-288-23	Sequence 23, Appl
39	17	8.3	743	4	US-09-282-357-5	Sequence 5, Appl
40	17	8.3	743	4	US-09-282-357-20	Sequence 20, Appl
41	17	8.3	743	4	US-09-282-357-23	Sequence 23, Appl
42	17	8.3	744	4	US-08-918-288-8	Sequence 8, Appl
43	17	8.3	744	4	US-09-282-357-8	Sequence 8, Appl
44	17	8.3	752	4	US-08-918-288-14	Sequence 14, Appl
45	17	8.3	752	4	US-08-918-288-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-083-521-3
; Sequence 3, Application US/09083521
; Patent No. 6048970
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,521
; FILING DATE: Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0527 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1213 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT10
; CLONE: 1691243
US-09-083-521-3

Query Match 14.7%; Score 30; DB 3; Length 1213;
Best Local Similarity 100.0%; Pred. No. 1.le-07;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 6 ttattacaccacaaattgtttgtctc 35

THIS PAGE BLANK (USPTO)

CC stimulating hormone (FSH). FSH is a heterodimeric gonadotropic hormone
CC secreted by the pituitary gland. It consists of two subunits, referred
CC to as alpha and beta. The protein is used to produce compositions
CC of the invention. The specification describes a stabilized dry powder
CC composition for delivery of a FSH to the deep lung of a mammalian
CC subject. The compositions promotes ovarian follicular development. The
CC composition is useful for treating female infertility.

XX
SQ Sequence 273 BP; 74 A; 73 C; 62 G; 64 T; 0 other;

Query Match 8.3%; Score 17; DB 21; Length 273;
Best Local Similarity 100.0%; Pred. NO. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 184 aggtcacagtaatggg 200
| | | | | | | | | |
Db 196 aggtcacagtaatggg 212

Search completed: September 20, 2002, 09:59:59
Job time: 11717 sec

PD 19-OCT-2000.
XX
PF 13-APR-2000; 2000WO-US09869.
XX
PR 13-APR-1999; 99US-0129121.
PR 20-APR-1999; 99US-0130099.
XX
PA (INHA-) INHALE THERAPEUTIC SYSTEMS INC.
PA (ELIL) LILLY & CO ELI.
XX
XX Nagarajan S, Patton JS, Bennett DB, Greene J, Chiang H;
PI Stults CLM, Venthoye G, Allen DL, Hughes BL, Stiff-Torvik M;
PI Wolff RK, Roeder WD;
XX
DR WPI; 2000-647398/62.
DR P-PSDB; AAB19375.
XX
XX Stabilized dry powder composition for delivery to the deep lung
PT comprising follicle-stimulating protein (FSP) and an excipient, useful
PT for the treatment of female infertility -
XX
XX Disclosure; Page 108; 125pp; English.
XX
XX The present sequence encodes an alpha subunit variant of a follicle
CC stimulating hormone (FSH). FSH is a heterodimeric gonadotropic hormone
CC secreted by the pituitary gland. It consists of two subunits, referred
CC to as alpha and beta. The protein is used to produce compositions
CC of the invention. The specification describes a stabilized dry powder
CC composition for delivery of a FSH to the deep lung of a mammalian
CC subject. The compositions promotes ovarian follicular development. The
CC composition is useful for treating female infertility.
XX
XX Sequence 267 BP; 73 A; 71 C; 61 G; 62 T; 0 other;
SQ

Query Match 8.3%; Score 17; DB 21; Length 267;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 184 agggtcacagtaatggg 200
|||||
Db 190 agggtcacagtaatggg 206
|||||

RESULT 14
AAC61708
ID AAC61708 standard; DNA; 270 BP.
XX
AC AAC61708;
XX
DT 06-MAR-2001 (first entry)
XX
DE DNA encoding an alpha subunit variant of a follicle stimulating hormone.
XX
KW Follicle stimulating hormone; FSH; gonadotropic hormone;
KW ovarian follicular development; infertility; ss.
XX
OS Homo sapiens.
XX
PN WO200061178-A1.
XX
PD 19-OCT-2000.
XX
PF 13-APR-2000; 2000WO-US09869.
XX
PR 13-APR-1999; 99US-0129121.
PR 20-APR-1999; 99US-0130099.
XX
PA (INHA-) INHALE THERAPEUTIC SYSTEMS INC.
PA (ELIL) LILLY & CO ELI.
XX
XX Nagarajan S, Patton JS, Bennett DB, Greene J, Chiang H;
PI Stults CLM, Venthoye G, Allen DL, Hughes BL, Stiff-Torvik M;

PI Wolff RK, Roeder WD;
XX
DR WPI; 2000-647398/62.
DR P-PSDB; AAB19374.
XX
XX Stabilized dry powder composition for delivery to the deep lung
PT comprising follicle-stimulating protein (FSP) and an excipient, useful
PT for the treatment of female infertility -
XX
XX Disclosure; Page 108; 125pp; English.
XX
XX The present sequence encodes an alpha subunit variant of a follicle
CC stimulating hormone (FSH). FSH is a heterodimeric gonadotropic hormone
CC secreted by the pituitary gland. It consists of two subunits, referred
CC to as alpha and beta. The protein is used to produce compositions
CC of the invention. The specification describes a stabilized dry powder
CC composition for delivery of a FSH to the deep lung of a mammalian
CC subject. The compositions promotes ovarian follicular development. The
CC composition is useful for treating female infertility.
XX
XX Sequence 270 BP; 74 A; 71 C; 62 G; 63 T; 0 other;
SQ

Query Match 8.3%; Score 17; DB 21; Length 270;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 184 agggtcacagtaatggg 200
|||||
Db 193 agggtcacagtaatggg 209
|||||

RESULT 15
AAC61707
ID AAC61707 standard; DNA; 273 BP.
XX
AC AAC61707;
XX
DT 06-MAR-2001 (first entry)
XX
DE DNA encoding an alpha subunit variant of a follicle stimulating hormone.
XX
KW Follicle stimulating hormone; FSH; gonadotropic hormone;
KW ovarian follicular development; infertility; ss.
XX
OS Homo sapiens.
XX
PN WO200061178-A1.
XX
PD 19-OCT-2000.
XX
PF 13-APR-2000; 2000WO-US09869.
XX
PR 13-APR-1999; 99US-0129121.
PR 20-APR-1999; 99US-0130099.
XX
PA (INHA-) INHALE THERAPEUTIC SYSTEMS INC.
PA (ELIL) LILLY & CO ELI.
XX
XX Nagarajan S, Patton JS, Bennett DB, Greene J, Chiang H;
PI Stults CLM, Venthoye G, Allen DL, Hughes BL, Stiff-Torvik M;
PI Wolff RK, Roeder WD;
XX
DR WPI; 2000-647398/62.
DR P-PSDB; AAB19373.
XX
XX Stabilized dry powder composition for delivery to the deep lung
PT comprising follicle-stimulating protein (FSP) and an excipient, useful
PT for the treatment of female infertility -
XX
XX Disclosure; Page 108; 125pp; English.
XX
XX The present sequence encodes an alpha subunit variant of a follicle

benign prostatic hyperplasia; acute prostatitis; testicular cancer; cryptorchidism; testicular disorder; proliferative disorder; lymphoma; leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer; liver cancer; lung cancer; cytostatic; ss.

Homo sapiens.

Key Location/Qualifiers
CDS 178..1650
/*tag= a
/product= "STMP1"

WO200172962-A2.
04-OCT-2001.
23-MAR-2001; 2001WO-US09410.
24-MAR-2000; 2000US-191929P.
(SAAT/) SAATCIOGLU F.
Saatioglu F;
WPI; 2001-662926/76.
P-PSDB; AAU10187.

New polynucleotide for the diagnosis, prevention and treatment for prostate and testis disorders, particularly prostate cancer, comprises prostate-specific or testis-specific nucleic acids -

Claim 5; Fig 4D; 114pp; English.

The invention relates to substantially pure prostate-specific or testis-specific polypeptides and the nucleic acids encoding them. Also included are vectors and host cells expressing the proteins, a transgenic animal expressing the protein, antibodies against the proteins, probes for detecting the nucleic acids, antisense molecules for the nucleic acids and methods of isolating modulators of the proteins. Compounds that modulate the prostate specific or testis specific polypeptide are useful to diagnose, prevent or treat disorders of the testis or prostate particularly prostate cancer, benign prostatic hyperplasia, acute prostatitis, testicular cancer, cryptorchidism, undescended, retractile, ascending or vanished testis. Other proliferative disorders for which the modulators may be used include lymphoma, leukaemia, melanoma, ovarian cancer, breast cancer, pancreatic cancer, liver cancer and lung cancer. The present sequence represents the open reading frame of a prostate specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.

Sequence 4329 BP; 1315 A; 817 C; 790 G; 1407 T; 0 other;

Query Match 14.7%; Score 30; DB 22; Length 4329;
Best Local Similarity 100.0%; Pred. No. 3.6e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ttatatacaccacaaactttgtcttgc 35
|||||
Db 1454 ttatatacaccacaaactttgtcttgc 1483

RESULT 12
AAS15800
ID AAS15800 standard; DNA; 2381 BP.
XX
AC AAS15800;
XX
XX 16-JAN-2002 (first entry)
XX Human Six-Transmembrane Protein of Prostate 1, STMP1, exon 6/3' UTR.
DE
XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;

benign prostatic hyperplasia; acute prostatitis; testicular cancer; cryptorchidism; testicular disorder; proliferative disorder; lymphoma; leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer; liver cancer; lung cancer; cytostatic; ds; exon 6.

Homo sapiens.

WO200172962-A2.
04-OCT-2001.
23-MAR-2001; 2001WO-US09410.
24-MAR-2000; 2000US-191929P.
(SAAT/) SAATCIOGLU F.
Saatioglu F;
WPI; 2001-662926/76.

New polynucleotide for the diagnosis, prevention and treatment for prostate and testis disorders, particularly prostate cancer, comprises prostate-specific or testis-specific nucleic acids -

Claim 6; Fig 4C; 114pp; English.

The invention relates to substantially pure prostate-specific or testis-specific polypeptides and the nucleic acids encoding them. Also included are vectors and host cells expressing the proteins, a transgenic animal expressing the protein, antibodies against the proteins, probes for detecting the nucleic acids, antisense molecules for the nucleic acids and methods of isolating modulators of the proteins. Compounds that modulate the prostate specific or testis specific polypeptide are useful to diagnose, prevent or treat disorders of the testis or prostate particularly prostate cancer, benign prostatic hyperplasia, acute prostatitis, testicular cancer, cryptorchidism, undescended, retractile, ascending or vanished testis. Other proliferative disorders for which the modulators may be used include lymphoma, leukaemia, melanoma, ovarian cancer, breast cancer, pancreatic cancer, liver cancer and lung cancer. The present sequence represents partial exon 6/3' UTR sequence of a prostate specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.

Sequence 2381 BP; 780 A; 415 C; 387 G; 799 T; 0 other;

Query Match 11.3%; Score 23; DB 22; Length 2381;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 ccccgagaggggtcacagtaag 198
|||||
Db 115 ccccgagaggggtcacagtaag 137

RESULT 13
AAC61709
ID AAC61709 standard; DNA; 267 BP.
XX
AC AAC61709;
XX
XX 06-MAR-2001 (first entry)
XX
DE DNA encoding an alpha subunit variant of a follicle stimulating hormone.
XX
XX Follicle stimulating hormone; FSH; gonadotropic hormone;
KW ovarian follicular development; infertility; ss.
XX
OS Homo sapiens.
XX
XX WO200061178-A1.
PN
XX

FT /product= "Human six transmembrane epithelial antigen
 FT of the prostate (STEAP)-2, alternative version"
 FT /note= "CDS does not include start and stop codon"
 FT /transl_except= (pos:1714..1722, aa:Asp-Ala)
 FT /transl_except= (pos:1834..1842, aa:Arg-Ser)
 FT /transl_except= (pos:1957..1965, aa:Glu-Gly)
 FT /transl_except= (pos:2050..2058, aa:Thr-Ser)
 FT /transl_except= (pos:2062..2070, aa:Asn-Phe)
 FT /note= "Inframe stop codon alters the reading frame"
 FT /partial
 XX
 PN WO200140276-A2.
 XX
 PD 07-JUN-2001.
 XX
 PF 06-DEC-2000; 2000WO-US33040.
 XX
 PR 06-DEC-1999; 99US-0455486.
 XX
 PA (UROG-) UROGENESYS INC.
 XX
 PI Afar DEH, Hubert RS, Raitano AB, Saffran DC, Mitchell SC, Faris M;
 PI Jakobovits A;
 XX
 DR WPI; 2001-367804/38.
 DR P-PSDB; AAE02781, AAE02841.
 XX
 PT New STEAP (six transmembrane epithelial antigen of the prostate)
 PT proteins, expressed in human cancers, useful for detecting and treating
 PT cancer -
 XX
 PS Claim 4; Fig 9A-9D; 187pp; English.
 XX
 CC The present sequence is human six transmembrane epithelial antigen of
 CC the prostate (STEAP)-2 clone GTD3 cDNA. STEAP is a member of cell
 CC surface serpentine transmembrane antigens. STEAP-2 gene is located on
 CC chromosome 7q21 and is used in gene therapy. Inhibiting the development
 CC or progression of a cancer (eg. prostate, colon, bladder, lung, ovarian
 CC and pancreatic) expressing STEAP or inhibiting growth or killing cells
 CC expressing STEAP in a patient, comprises administering a vaccine
 CC composition to the patient. Treating a patient with a cancer that
 CC expresses STEAP, or inhibiting growth or killing cells expressing STEAP,
 CC comprises administering to the patient a vector encoding single chain
 CC monoclonal antibody that comprises the variable domains of the heavy and
 CC light chains of the monoclonal antibody that specifically binds to STEAP,
 CC such that the vector delivers the single chain monoclonal antibody coding
 CC sequence to the cancer cells and the encoded single chain monoclonal
 CC antibody is expressed intracellularly.
 CC Note: The present sequence is also shown in sequence listing of the
 CC specification, but it lacks nucleotides at its 5' end.
 XX
 SQ Sequence 2453 BP; 648 A; 537 C; 520 G; 748 T; 0 other;

Query Match 14.7%; Score 30; DB 22; Length 2453;
 Best Local Similarity 100.0%; Pred. No. 3.6e-06;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 ttatatacaccacaaacttggtctgctc 35
 |||
 Db 1631 ttatatacaccacaaacttggtctgctc 1660
 RESULT 10
 AAS64300
 ID AAS64300 standard; cDNA; 3900 BP.
 XX
 AC AAS64300;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #104.
 XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR P-PSDB; ABG00113.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 104; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 3900 BP; 1161 A; 798 C; 892 G; 1049 T; 0 other;

Query Match 14.7%; Score 30; DB 23; Length 3900;
 Best Local Similarity 100.0%; Pred. No. 3.6e-06;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 ttatatacaccacaaacttggtctgctc 35
 |||
 Db 1406 ttatatacaccacaaacttggtctgctc 1435
 RESULT 11
 AAS15801
 ID AAS15801 standard; cDNA; 4329 BP.
 XX
 AC AAS15801;
 XX
 DT 16-JAN-2002 (first entry)
 XX
 DE Human ORF of Six-Transmembrane Protein of Prostate 1, STMP1.
 XX
 DE Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
 XX

PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG12306.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
FT biodiversity
XX
XX Claim 1; SEQ ID No 12297; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS4197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2192 BP; 559 A; 507 C; 551 G; 575 T; 0 other;

Query Match 14.7%; Score 30; DB 23; Length 2192;
Best Local Similarity 100.0%; Pred. No. 3.6e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ttatatcacaccacaaactttgtttctgctc 35
|||||
Db 1406 ttatatcacaccacaaactttgtttctgctc 1435

RESULT 8
AAS15810
ID AAS15810 standard; cDNA; 2238 BP.
XX
XX AAS15810;

DT 16-JAN-2002 (first entry)

XX Human ORF2 of Six-Transmembrane Protein of Prostate 1, STMP1.

XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ss; ORF2.

XX Homo sapiens.

XX Key Location/Qualifiers
FH CDS 188..1552
FT /*tag= a
FT /product= "STMP1, ORF2"

PN W0200172962-A2.

XX 04-OCT-2001.

XX

PF 23-MAR-2001; 2001WO-US09410.

XX
XX 24-MAR-2000; 2000US-191929P.

XX (SAAT/) SAATCIOGLU F.

XX Saatioglu F;

XX WPI; 2001-662926/76.

DR P-PSDB; AAU10188.

XX New polynucleotide for the diagnosis, prevention and treatment for

PT prostate and testis disorders, particularly prostate cancer, comprises

XX prostate-specific or testis-specific nucleic acids

XX Claim 5; Fig 4G; 114pp; English.

XX The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence represents the second open reading frame of a prostate
CC specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.

XX
SQ Sequence 2238 BP; 607 A; 457 C; 453 G; 721 T; 0 other;

Query Match 14.7%; Score 30; DB 22; Length 2238;
Best Local Similarity 100.0%; Pred. No. 3.6e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ttatatcacaccacaaactttgtttctgctc 35
|||||
Db 1464 ttatatcacaccacaaactttgtttctgctc 1493

RESULT 9
AAD07072
ID AAD07072 standard; cDNA; 2453 BP.
XX
XX AAD07072;

DT 06-AUG-2001 (first entry)

XX Human six transmembrane epithelial antigen of prostate-2 clone GTD3 CDNA.

XX Human; cytostatic; antiproliferative; vaccine; gene therapy;
KW six transmembrane epithelial antigen of the prostate-2; STAP-2;
KW chromosome 7q21; cancer; prostate; colon; bladder; lung; ovarian;
KW pancreatic; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
FH RBS 352..358
FT /*tag= a
FT /note= "Kozak region"

FT CDS 355..1719
FT /*tag= b
FT /product= "Human six transmembrane epithelial antigen
FT of the prostate (STAP)-2"
FT 709..2073
FT /*tag= c

```

XX PS Claim 4; Fig 4E; 114pp; English.
XX
CC The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence encodes a prostate specific protein, Six-Transmembrane
CC Protein of Prostate 1, STMP1.
XX
SQ Sequence 1680 BP; 467 A; 334 C; 373 G; 506 T; 0 other;

Query Match          14.7%; Score 30; DB 22; Length 1680;
Best Local Similarity 100.0%; Pred. No. 3.6e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 ttatatacaccacaaacttgttctgtc 35
Db 1454 ttatatacaccacaaacttgttctgtc 1483

RESULT 6
AAS15793
ID AAS15793 standard; DNA; 1725 BP.
AC AAS15793;
XX
DT 16-JAN-2002 (first entry)
DE Human DNA for Six-Transmembrane Protein of Prostate 1, STMP1.
XX
KW Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_feature 89
FT /*tag= a
FT /*note= "Represents 338 nucleotides of intron 1"
FT 162
FT /*tag= b
FT /*note= "Represents 12713 nucleotides of intron 2"
FT 200..1702
FT /*tag= c
FT /*product= "STMP1"
FT 697
FT misc_feature
FT /*tag= d
FT /*note= "Represents 1396 nucleotides of intron 3"
FT 1225
FT misc_feature
FT /*tag= e
FT /*note= "Represents 2372 nucleotides of intron 4"
FT 1410
FT misc_feature
FT /*tag= f
FT /*note= "Represents 2299 nucleotides of intron 5"
XX
XX WO200172962-A2.
XX
PD 04-OCT-2001.

```

```

XX PF 23-MAR-2001; 2001WO-US09410.
XX
XX PR 24-MAR-2000; 2000US-191929P.
XX
XX PA (SAAT/) SAATCIOGLU F.
XX
XX PI Saatcioglu F;
XX
XX DR WPI; 2001-662926/76.
XX DR P-PSDB; AAU10187.
XX
XX PT New polynucleotide for the diagnosis, prevention and treatment for
XX prostate and testis disorders, particularly prostate cancer, comprises
XX prostate-specific or testis-specific nucleic acids
XX
XX PS Example 3; Fig 4B; 114pp; English.
XX
XX CC The invention relates to substantially pure prostate-specific or
XX testis-specific polypeptides and the nucleic acids encoding them.
XX Also included are vectors and host cells expressing the proteins, a
XX transgenic animal expressing the protein, antibodies against the
XX proteins, probes for detecting the nucleic acids, antisense molecules
XX for the nucleic acids and methods of isolating modulators of the
XX proteins. Compounds that modulate the prostate specific or testis
XX specific polypeptide are useful to diagnose, prevent or treat disorders
XX of the testis or prostate particularly prostate cancer, benign
XX prostatic hyperplasia, acute prostatitis, testicular cancer,
XX cryptorchidism, undescended, retractile, ascending or vanished
XX testis. Other proliferative disorders for which the modulators may be
XX used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
XX cancer, pancreatic cancer, liver cancer and lung cancer. The
XX present sequence encodes a prostate specific protein, Six-Transmembrane
XX Protein of Prostate 1, STMP1.
XX
XX SQ Sequence 1725 BP; 476 A; 340 C; 387 G; 517 T; 5 other;

Query Match          14.7%; Score 30; DB 22; Length 1725;
Best Local Similarity 100.0%; Pred. No. 3.6e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 ttatatacaccacaaacttgttctgtc 35
Db 1506 ttatatacaccacaaacttgttctgtc 1535

RESULT 7
AAS76493
ID AAS76493 standard; cDNA; 2192 BP.
AC AAS76493;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE DNA encoding novel human diagnostic protein #12297.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX PA (HYSE-) HYSEQ INC.
XX

```

Claim 7; Page 67; 72pp: English.

PS This sequence represents cDNA encoding human prostate growth-associated
 XX protein PGAMP-1. Nucleotides encoding PGAMP-1 were initially identified
 CC in a prostate cDNA library, this sequence representing a consensus.
 CC Human prostate growth associated membrane proteins PGAMP-1 and PGAMP-2
 CC (AA52590) may be used to raise specific antibodies and to screen for
 CC specific modulators (agonists, antagonists or other potential
 CC therapeutic agents). Antagonists of PGAMP are used to treat or prevent a
 CC wide range of cancers (solid tumours, leukaemia, lymphoma etc.) and
 CC reproductive disorders (such as infertility, endometriosis, polycystic
 CC ovarian syndrome, prostatitis). PGAMP-encoding nucleic acids, its
 CC fragments and complements, may be used for recombinant production of
 CC PGAMP proteins, in gene therapy (e.g., as antisense molecules, triplex-
 CC forming molecules and ribozymes), and as diagnostic probes and primers.
 CC Anti-PGAMP antibodies may be used for diagnosis and monitoring of
 CC PGAMP-related diseases by standard immunoassays, as therapeutic
 CC antagonists (including targeted delivery of other drugs), and in
 CC competitive drug screens.
 XX
 SQ Sequence 1213 BP; 335 A; 239 C; 215 G; 424 T; 0 other;

Query Match 14.7%; Score 30; DB 21; Length 1213;
 Best Local Similarity 100.0%; Pred. No. 3.6e-06;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ttatatacaccacaaacttggtctgtctc 35
 |||||
 Db 420 ttatatacaccacaaacttggtctgtctc 449

RESULT 4

AAS15811
 ID AAS15811 standard; cDNA; 1561 BP.

AC AAS15811;

XX 16-JAN-2002 (first entry)

DE Human cDNA encoding ORF2 of Six-Transmembrane Protein of Prostate 1.

XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
 KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
 KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
 KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
 KW liver cancer; lung cancer; cytostatic; ss; ORF2.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH 188..1552
 FT CDS /*tag= a /product= "STMP1, ORF2"

FT WO200172962-A2.

XX 04-OCT-2001.

XX 23-MAR-2001; 2001WO-US09410.

XX 24-MAR-2000; 2000US-191929P.

XX (SAAT/) SAATCIOGLU F.

XX Saatcioglu F;

XX WPI; 2001-662926/76.

XX P-PSDB; AAU10188.

XX New polynucleotide for the diagnosis, prevention and treatment for

PT prostate and testis disorders, particularly prostate cancer, comprises
 PT prostate-specific or testis-specific nucleic acids -

XX Claim 4; Fig 4H; 114pp: English.

PS The invention relates to substantially pure prostate-specific or
 XX testis-specific polypeptides and the nucleic acids encoding them.
 CC Also included are vectors and host cells expressing the proteins, a
 CC transgenic animal expressing the protein, antibodies against the
 CC proteins, probes for detecting the nucleic acids, antisense molecules
 CC for the nucleic acids and methods of isolating modulators of the
 CC proteins. Compounds that modulate the prostate specific or testis
 CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,
 CC cryptorchidism, undescended, retractile, ascending or vanished
 CC testis. Other proliferative disorders for which the modulators may be
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The
 CC present sequence encodes prostate specific protein, Six-Transmembrane
 CC Protein of Prostate 1, STMP1, ORF2.
 XX
 SQ Sequence 1561 BP; 420 A; 319 C; 346 G; 476 T; 0 other;

Query Match 14.7%; Score 30; DB 22; Length 1561;
 Best Local Similarity 100.0%; Pred. No. 3.6e-06;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ttatatacaccacaaacttggtctgtctc 35
 |||||
 Db 1464 ttatatacaccacaaacttggtctgtctc 1493

RESULT 5

AAS15802
 ID AAS15802 standard; cDNA; 1680 BP.

AC AAS15802;

XX 16-JAN-2002 (first entry)

DE Human cDNA encoding Six-Transmembrane Protein of Prostate 1, STMP1.

XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
 KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
 KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
 KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
 KW liver cancer; lung cancer; cytostatic; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH 178..1650
 FT CDS /*tag= a /product= "STMP1"

FT WO200172962-A2.

XX 04-OCT-2001.

XX 23-MAR-2001; 2001WO-US09410.

XX 24-MAR-2000; 2000US-191929P.

XX (SAAT/) SAATCIOGLU F.

XX Saatcioglu F;

XX WPI; 2001-662926/76.

XX P-PSDB; AAU10187.

XX New polynucleotide for the diagnosis, prevention and treatment for

PT prostate and testis disorders, particularly prostate cancer, comprises
 PT prostate-specific or testis-specific nucleic acids -

PT New polynucleotide for the diagnosis, prevention and treatment for
 PT prostate and testis disorders, particularly prostate cancer, comprises
 PS prostate-specific or testis-specific nucleic acids -
 PS Claim 6; Fig 4C; 114pp; English.
 XX
 XX The invention relates to substantially pure prostate-specific or
 CC testis-specific polypeptides and the nucleic acids encoding them.
 CC Also included are vectors and host cells expressing the proteins, a
 CC transgenic animal expressing the protein, antibodies against the
 CC proteins, probes for detecting the nucleic acids, antisense molecules
 CC for the nucleic acids and methods of isolating modulators of the
 CC proteins. Compounds that modulate the prostate specific or testis
 CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,
 CC cryptorchidism, undescended, retractile, ascending or vanished
 CC testis. Other proliferative disorders for which the modulators may be
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The
 CC present sequence represents partial exon 6 sequence of a prostate
 CC specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.
 XX
 SQ Sequence 148 BP; 36 A; 29 C; 26 G; 57 T; 0 other;

Query Match 14.7%; Score 30; DB 22; Length 148;
 Best Local Similarity 100.0%; Pred. No. 3.5e-06;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ttatataccaccacaaacttggttgtgctc 35
 |||||
 Db 92 ttatataccaccacaaacttggttgtgctc 121

RESULT 2
 AAS15808
 ID AAS15808 standard; DNA; 148 BP.

AC AAS15808;
 DT 16-JAN-2002 (first entry)
 DE Human Six-Transmembrane Protein of Prostate 1, STMP1, ORF2 exon 6.
 XX
 XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
 KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
 KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
 KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
 KW liver cancer; lung cancer; cystostatic; ds; exon 6; ORF2.
 XX
 OS Homo sapiens.
 XX
 XX WO200172962-A2.
 PN
 XX
 PD 04-OCT-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09410.
 PF
 XX
 XX 24-MAR-2000; 2000US-191929P.
 PR
 XX
 XX (SAAT/) SAATCIOGLU F.
 PA
 XX
 XX Saatcioglu F;
 PI
 XX
 XX WPI; 2001-662926/76.
 DR
 XX
 XX New polynucleotide for the diagnosis, prevention and treatment for
 PT prostate and testis disorders, particularly prostate cancer, comprises
 PT prostate-specific or testis-specific nucleic acids -
 XX
 XX Claim 6; Fig 4F; 114pp; English.

CC The invention relates to substantially pure prostate-specific or
 CC testis-specific polypeptides and the nucleic acids encoding them.
 CC Also included are vectors and host cells expressing the proteins, a
 CC transgenic animal expressing the protein, antibodies against the
 CC proteins, probes for detecting the nucleic acids, antisense molecules
 CC for the nucleic acids and methods of isolating modulators of the
 CC proteins. Compounds that modulate the prostate specific or testis
 CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,
 CC cryptorchidism, undescended, retractile, ascending or vanished
 CC testis. Other proliferative disorders for which the modulators may be
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The
 CC present sequence represents exon 6 of a prostate specific protein,
 CC Six-Transmembrane Protein of Prostate 1, STMP1, alternatively
 XX spliced version, ORF2.

SQ Sequence 148 BP; 36 A; 29 C; 26 G; 57 T; 0 other;

Query Match 14.7%; Score 30; DB 22; Length 148;
 Best Local Similarity 100.0%; Pred. No. 3.5e-06;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ttatataccaccacaaacttggttgtgctc 35
 |||||
 Db 92 ttatataccaccacaaacttggttgtgctc 121

RESULT 3
 AAZ46296
 ID AAZ46296 standard; cDNA; 1213 BP.

XX AAZ46296;
 XX
 XX 07-MAR-2000 (first entry)
 DT
 XX
 XX DE
 XX Human prostate growth-associated membrane protein PGAMP-1 cDNA.
 XX
 XX Prostate growth-associated membrane protein; PGAMP-1; prostate;
 KW consensus; antibody; screening; modulator; agonist; antagonist;
 KW therapeutic agent; cancer; solid tumour; leukaemia; lymphoma;
 KW reproductive disorder; infertility; endometriosis;
 KW polycystic ovarian syndrome; prostatitis; recombinant expression;
 KW gene therapy; antisense therapy; ribozyme; diagnosis; diagnosis;
 KW monitoring; immunoassay; targeting; drug delivery; drug screening; ds.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH 83..508
 FT CDS /*tag= a
 FT /*product= "Human PGAMP-1"
 FT
 XX WO9961469-A2.
 XX
 XX 02-DEC-1999.
 PD
 XX
 XX 17-MAY-1999; 99WO-US10888.
 PF
 XX
 XX 22-MAY-1998; 98US-0083521.
 PR
 XX
 XX (INCY-) INCYTE PHARM INC.
 PA
 XX
 XX Lal P, Guegler KJ, Corley NC;
 PI
 XX
 XX WPI; 2000-062671/05.
 DR
 XX P-PSDB; AAY52589.
 DR
 XX New human prostate growth-associated membrane proteins, for treating or
 PT preventing cancer and reproductive disorders -
 PT

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:59:56 ; Search time 600.75 seconds
(without alignments)
583.022 Million cell updates/sec

Title: US-09-802-520-8

Perfect score: 204

Sequence: 1 cagagttttatcacaccacaa.....gggtcacagtaatggatga 204

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	14.7	148	22	AA515799
2	30	14.7	148	22	AA515808
3	30	14.7	1213	21	AA246296
4	30	14.7	1561	22	AA515811
5	30	14.7	1680	22	AA515802
6	30	14.7	1725	22	AA515793
7	30	14.7	2192	23	AA576493
8	30	14.7	2238	22	AA515810
9	30	14.7	2453	22	AA007072

10	30	14.7	3900	23	AAS64300	DNA encoding novel
11	30	14.7	4329	22	AAS15801	Human ORF of Six-T
12	23	11.3	2381	22	AAS15800	Human Six-Transmem
13	17	8.3	287	21	AAC61709	DNA encoding an al
14	17	8.3	270	21	AAC61708	DNA encoding an al
15	17	8.3	273	21	AAC61707	DNA encoding an al
16	17	8.3	276	21	AAC61685	DNA encoding the a
17	17	8.3	276	21	AAC61686	DNA encoding the a
18	17	8.3	276	21	AAZ45449	Nucleotide sequenc
19	17	8.3	276	21	AAZ45450	Nucleotide sequenc
20	17	8.3	312	16	AAT03237	Gonadotropin alpha
21	17	8.3	312	22	AAD08805	Human single chain
22	17	8.3	317	22	AAS08505	DNA encoding gonad
23	17	8.3	317	22	AAD08806	Human single chain
24	17	8.3	348	22	AAH46590	Human anterior pit
25	17	8.3	351	18	AAV02211	Secreted protein h
26	17	8.3	351	21	AAAS3565	Human chorionic go
27	17	8.3	351	24	AA517402	Human cDNA encodin
28	17	8.3	444	21	AAC00111	Human secreted pro
29	17	8.3	456	22	AAH65154	C glutamic codin
30	17	8.3	592	12	AAQ10075	Engineered human a
31	17	8.3	592	17	AAT41695	Dimeric glycoprote
32	17	8.3	631	20	AAZ31747	Human chorionic go
33	17	8.3	681	22	AAD08804	Human single chain
34	17	8.3	683	21	AAF13416	Aspergillus oryzae
35	17	8.3	693	22	AAD08802	Human single chain
36	17	8.3	702	22	AAD08792	Human single chain
37	17	8.3	707	16	AAT03236	Single chain gonad
38	17	8.3	707	22	AAS08503	DNA encoding singl
39	17	8.3	707	22	AAD08803	Human single chain
40	17	8.3	717	22	AAD08788	Human single chain
41	17	8.3	717	22	AAD08798	Human single chain
42	17	8.3	717	22	AAD08800	Human single chain
43	17	8.3	718	22	AAD08790	Human single chain
44	17	8.3	719	16	AAT03235	Single chain gonad
45	17	8.3	719	22	AAS08501	DNA encoding singl

ALIGNMENTS

RESULT 1

AA515799

ID AA515799 standard; DNA; 148 BP.

XX AC AA515799;

XX AC AA515799;

DT 16-JAN-2002 (first entry)

XX Human Six-Transmembrane Protein of Prostate 1, STMP1, partial exon 6.

XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer; benign prostatic hyperplasia; acute prostatitis; testicular cancer; cryptorchidism; testicular disorder; proliferative disorder; lymphoma; leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer; liver cancer; lung cancer; cytostatic; ds; exon 6.

XX Homo sapiens.

XX WO200172962-A2.

XX 04-OCT-2001.

XX 23-MAR-2001; 2001WO-US09410.

XX 24-MAR-2000; 2000US-191929P.

XX (SAAT/) SAATC1OGLU F.

XX Saatcioglu F;

XX WPI; 2001-662926/76.

XX

THIS PAGE BLANK (USPTO)

MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

BAC ends sequenced at TIGR from the RPC111 BAC library. Designed
and developed at the Stanford Human Genome Center.

FEATURES

source

Location/Qualifiers

1..629
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="5"
/clone_lib="Human"

STS

primer_bind

183..482

primer_bind complement(460..482)

194 a 99 c 127 g 209 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 9.3%; Score 19; DB 11; Length 629;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 112 aagaaaggctgggaaaga 130

|||||

Db 556 AAGAAAGGCTGGGAAAGA 574

Search completed: September 20, 2002, 09:57:55
Job time: 13333 sec


```

Query Match      9.8%; Score 20; DB 2; Length 205085;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 cgaagctaaacgaattaa 113
|||||
Db 196162 CGAAGCTAAACGAATTAA 196181

RESULT 14
AC092404/c
LOCUS      227144 bp      DNA      linear      HTG 04-JUL-2001
DEFINITION Mus musculus chromosome 5 clone RP23-261D18 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 3 unordered pieces.
ACCESSION  AC092404
VERSION     AC092404.1 GI:14595778
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
ORGANISM    Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
             1 (bases 1 to 227144)
             Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
             Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L.,
             Granite,S., Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E.,
             Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B.,
             Masiello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
             Pearson,R., Prasad,A., Shevchenko,Y., Snyder,B., Stantirpop,S.,
             Thomas,J.W., Thomas,F.J., Touchman,J.W., Tsurgueon,C., Vogt,J.L.,
             Walker,M.A., Weherby,K.D., Zhang,L.-H. and Green,E.D.
             NISC Comparative Sequencing Initiative
             Unpublished
             2 (bases 1 to 227144)
             Green,E.D.
             Direct Submission
             Submitted (04-JUL-2001) NIH Intramural Sequencing Center, 8717
             Government Circle, Gaithersburg, MD 20877, USA
             ----- Genome Center
             Center: NIH Intramural Sequencing Center
             Center code: NISC
             Web site: http://www.nisc.nih.gov
             Contact: nisc_mouse@hgri.nih.gov
             ----- Project Information
             Center project name: qv
             Center clone name: 261D18
             ----- Summary Statistics
             Sequencing vector: plasmid; n/a; 100% of reads
             Chemistry: Dye-terminator Big Dye; 100% of reads
             Assembly program: Phrap; version 0.990319
             Consensus quality: 225678 bases at least Q40
             Consensus quality: 225959 bases at least Q30
             Consensus quality: 226087 bases at least Q20
             Insert size: 206000; agarose-fp
             Insert size: 225000; pulse-field-gel
             Insert size: 226944; sum-of-contigs
             Quality coverage: 14.02x in Q20 bases; agarose-fp
             Quality coverage: 12.84x in Q20 bases; pulse-field-gel
             Quality coverage: 12.73x in Q20 bases; sum-of-contigs
             -----
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 3 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence.
             * as soon as it is available and the accession number will
             * be preserved.
             * 1 26708: contig of 26708 bp in length
             * 26808: gap of unknown length
             * 26809 109655: contig of 82847 bp in length
             * 109656 109755: gap of unknown length
FEATURES
source
Location/Qualifiers
1..227144
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP23-261D18"
/clone_lib="RPC1 mouse BAC library 23"
1..26708
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
26809..109655
/note="assembly_fragment"
clone_end:T7
vector_side:left"
109756..227144
/note="assembly_fragment"
BASE COUNT 68114 a 46378 c 46791 g 65638 t 223 others
ORIGIN

Query Match      9.8%; Score 20; DB 2; Length 227144;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 cgaagctaaacgaattaa 113
|||||
Db 118681 CGAAGCTAAACGAATTAA 118662

RESULT 15
G57411
LOCUS      G57411      629 bp      DNA      linear      STS 30-MAR-2000
DEFINITION SHGC-103088 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION  G57411
VERSION     G57411.1 GI:6122580
KEYWORDS    STS.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
             1 (bases 1 to 629)
             Olivier,M. and Cox,D.R.
             TITLE
             Unpublished, Olivier, M., Cox, D.R. (2000)
             JOURNAL
             Unpublished
             COMMENT
             Contact: Michael Olivier, David R. Cox
             Stanford Human Genome Center
             Stanford University School of Medicine
             4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
             Tel: (650) 320-5800
             Fax: (650) 320-5801
             Email: olivier@shgc.stanford.edu
             Primer A: AGTTGGAGTGACGATTAACCGA
             Primer B: TTCAGGGTGAACCTTCTAGCAC
             STS size: 300
             PCR Profile:
             Initial incubation: 95 degrees C for 10 minutes
             Denaturation: 94 degrees C for 30 seconds
             Annealing: 60 degrees C for 30 seconds
             Polymerization: 72 degrees C for 23 seconds
             PCR Cycles: 30
             Thermal Cycler: Perkin Elmer 9700
             Protocol:
             Template: 25 ng
             Primer: each 1 uM
             dNTPs: each 200 uM
             AmpliTaq Gold Polymerase: 0.07 units/ul
             Total Vol: 5 ul
             Buffer:

```

```

/chromosome="15"
/map="15ql15"
/clone="RP11-292p13"
/clone_lib="RPC1 human BAC library 11"
/note="This clone overlaps CTD-2165P21 AC025270,
RP11-302D8 AC072032 and RP11-456H13 AC091245. Data from
overlapping BACs were added and the consensus sequence
determined from RP11-292P13 to the extent possible."
misc_feature
1..41260
/note="overlap with CTD-2165P21, AC025270"
misc_feature
58018..199503
/note="overlap with RP11-456H13, AC091245"
72240..72280
/note="overlap with RP11-302D8, AC072032"
misc_feature
102297..199503
/note="low quality data"
unsure
118860..118876
/note="overlap with RP11-302D8, AC072032"
/note="low quality data"
unsure
118943..118945
/note="low quality data"
BASE COUNT 57713 a 42859 c 41648 g 57283 t
ORIGIN
Query Match 9.8%; Score 20; DB 9; Length 199503;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 112 aagaaagctgggaaag 131
|||||
Db 68341 AAGAAAGCTGGGGAAG 68360

RESULT 13
AC026813 205085 bp DNA linear HTG 27-OCT-2001
LOCUS Mus musculus chromosome 5 clone RP23-119M19 strain C57BL6/J,
DEFINITION WORKING DRAFT SEQUENCE, 10 unordered pieces.
ACCESSION AC026813
VERSION AC026813.2 GI:16506398
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 205085)
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-Q.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Masiello, C., Mastrian, S.D.,
McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y.,
Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 205085)
Green, E.D.
Direct Submission
Submitted (24-MAR-2000) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
On Oct 27, 2001 this sequence version replaced gi:7321458.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nih.gov
----- Project Information
Center project name: xh
Center clone name: 119M19
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 199516 bases at least Q40
Consensus quality: 200596 bases at least Q30
Consensus quality: 201000 bases at least Q20
Insert size: 197000; agarose-fp
Insert size: 192000; pulse-field-gel
Insert size: 204185; sum-of-contigs
Quality coverage: 9.36x in Q20 bases; agarose-fp
Quality coverage: 9.60x in Q20 bases; pulse-field-gel
Quality coverage: 9.03x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4428: contig of 4428 bp in length
* 4429: gap of unknown length
* 4529: contig of 5025 bp in length
* 9553: gap of unknown length
* 22746: contig of 13093 bp in length
* 22846: gap of unknown length
* 33954: contig of 11108 bp in length
* 34054: gap of unknown length
* 47831: contig of 13777 bp in length
* 47931: gap of unknown length
* 62010: contig of 14079 bp in length
* 62011: gap of unknown length
* 81789: contig of 19679 bp in length
* 81790: gap of unknown length
* 81890: contig of 28030 bp in length
* 109920: gap of unknown length
* 110020: contig of 43216 bp in length
* 153235: gap of unknown length
* 153336: contig of 51750 bp in length.
* 205085: contig of 51750 bp in length.
FEATURES
Location/Qualifiers
1..205085
/organism="Mus musculus"
/db_xref="C57BL6/J"
/chromosome="5"
/clone="rp23-119M19"
/clone_lib="RPC1 mouse BAC library 23"
1..4428
/note="assembly_fragment"
4529..9553
/note="assembly_fragment"
9654..22746
/note="assembly_fragment"
22847..33954
/note="assembly_fragment"
34055..47831
/note="assembly_fragment"
clone_end:T7
vector_side:right"
47932..62010
/note="assembly_fragment"
62111..81789
/note="assembly_fragment"
81890..109919
/note="assembly_fragment"
110020..153235
/note="assembly_fragment"
153336..205085
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
BASE COUNT 61131 a 40716 c 41296 g 61026 t 916 others
ORIGIN

```

```
RESULT 11
AC091245 194311 bp DNA linear PRI 27-OCT-2001
LOCUS Homo sapiens chromosome 15 clone RP11-456H13 map 15q15, complete
DEFINITION sequence.
ACCESSION AC091245
VERSION AC091245.2 GI:16506376
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 194311)
AUTHORS Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D., and Hood,L.
Pate,D. and Hood,L.
Direct Submission
Submitted (10-APR-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
3 (bases 1 to 194311)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
Direct Submission
Submitted (27-OCT-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
On Oct 27, 2001 this sequence version replaced gi:13569969.
----- Genome Center
Center: Multimegabase Sequencing Center
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Note: Data from overlapping clones AC018901 [Drafting center:
UWMSCL, AC072032 [Drafting center: BCM] and AC090888 [Drafting
center: WTBRL] were added for finishing.
Location/Qualifiers
source 1. 194311
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q15"
/clone="RP11-456H13"
/clone_lib="RPCI human BAC library 11"
/note="This clone overlaps RP11-292P13 [AC018901 and
AC07828], RP11-302D8 AC072032 and CTD-2014N11 AC090888.
Data from overlapping BACs were added and the consensus
sequence determined from RP11-456H13 to the extent
possible."
1. 132638
/note="overlap with RP11-292P13, AC018901"
14140. .14193
/note="low quality data."
misc_feature 44246. .194311
/note="Overlap with RP11-302D8, AC072032"
150611. .150613
/note="low quality data"

Query Match 9.8%; Score 20; DB 9; Length 194311;
Best Local Similarity 100.0%; Pred. No. 2.8; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

QY 112 aagaagcgtgggaaag 131
|||||
Db 10282 AAGAAAGCTGGGAAGAG 10301
|||||

RESULT 12
AC018901 199503 bp DNA linear PRI 12-SEP-2001
LOCUS Homo sapiens chromosome 15 clone RP11-292P13 map 15q15, complete
DEFINITION sequence.
ACCESSION AC018901
VERSION AC018901.8 GI:15559174
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 199503)
AUTHORS Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
Pate,D. and Hood,L.
Sequencing of human chromosome 15 D15S146-D15S117 region
Unpublished
2 (bases 1 to 199503)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
and Hood,L.
Direct Submission
Submitted (22-DEC-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
3 (bases 1 to 199503)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
Direct Submission
Submitted (12-SEP-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
On Sep 12, 2001 this sequence version replaced gi:13569966.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UWMSCL
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Note: Data from overlapping clones AC025270 [Drafting center:
UWMSCL, AC072032 [Drafting center: BCM] and AC091245 [Drafting
center: UWMSCL] were added for finishing.
Location/Qualifiers
source 1. 199503
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source Note: Data from overlapping clones AC025270 [Drafting center:
UWMSCL, AC072032 [Drafting center: BCM] and AC091245 [Drafting
center: UWMSCL] were added for finishing.
Location/Qualifiers
1. 199503
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 88: contig of 88 bp in length
* 89 188: gap of 100 bp
* 189 1247: contig of 1059 bp in length
* 1248 1347: gap of 100 bp
* 1348 1956: contig of 609 bp in length
* 1957 2056: gap of 100 bp
* 2057 3554: contig of 1498 bp in length
* 3555 3654: gap of 100 bp
* 3655 4746: contig of 1092 bp in length
* 4747 4846: gap of 100 bp
* 4847 6735: contig of 1889 bp in length
* 6736 6835: gap of 100 bp
* 6836 8018: contig of 1183 bp in length
* 8019 8118: gap of 100 bp
* 8119 9450: contig of 1332 bp in length
* 9451 9550: gap of 100 bp
* 9551 10787: contig of 1237 bp in length
* 10788 12735: contig of 1848 bp in length
* 12736 12835: gap of 100 bp
* 12836 14917: contig of 2082 bp in length
* 14918 15017: gap of 100 bp
* 15018 16744: contig of 1727 bp in length
* 16745 16844: gap of 100 bp
* 16845 19913: contig of 3069 bp in length
* 19914 20013: gap of 100 bp
* 20014 32329: contig of 3226 bp in length
* 32340 23339: gap of 100 bp
* 23340 26001: contig of 2662 bp in length
* 26002 26101: gap of 100 bp
* 26102 30491: contig of 4390 bp in length
* 30492 30591: gap of 100 bp
* 30592 33976: contig of 3385 bp in length
* 33977 34076: gap of 100 bp
* 34077 38166: contig of 4090 bp in length
* 38167 38266: gap of 100 bp
* 38267 43587: contig of 5321 bp in length
* 43588 43687: gap of 100 bp
* 43688 50548: contig of 6861 bp in length
* 50549 50648: gap of 100 bp
* 50649 60679: contig of 10031 bp in length
* 60680 60779: gap of 100 bp
* 60780 68485: contig of 7706 bp in length
* 68486 68585: gap of 100 bp
* 68586 78448: contig of 9863 bp in length
* 78449 78548: gap of 100 bp
* 78549 85993: contig of 7445 bp in length
* 85994 86093: gap of 100 bp
* 86094 94636: contig of 8543 bp in length
* 94637 94736: gap of 100 bp
* 94737 114104: contig of 19368 bp in length
* 114105 114204: gap of 100 bp
* 114205 142862: contig of 28658 bp in length
* 142863 142962: gap of 100 bp
* 142963 167907: contig of 24945 bp in length
* 167908 168007: gap of 100 bp
* 168008 186406: contig of 18399 bp in length.

FEATURES
source

1. .186406
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-292P13"

misc_feature
1. .88
/clone_lib="RPC1-11 Human Male BAC"
/note="assembly_fragment"
clone_end:SP6
vector_side:left
189. .1247
/note="assembly_fragment"
1348. .1956
/note="assembly_fragment"
2057. .3554
/note="assembly_fragment"
3655. .4746
/note="assembly_fragment"
4847. .6735
/note="assembly_fragment"
6836. .8018
/note="assembly_fragment"
8119. .9450
/note="assembly_fragment"
9551. .10787
/note="assembly_fragment"
10888. .12735
/note="assembly_fragment"
12836. .14917
/note="assembly_fragment"
15018. .16744
/note="assembly_fragment"
16845. .19913
/note="assembly_fragment"
20014. .23239
/note="assembly_fragment"
23340. .26001
/note="assembly_fragment"
26102. .30491
/note="assembly_fragment"
30592. .33976
/note="assembly_fragment"
34077. .38166
/note="assembly_fragment"
38267. .43587
/note="assembly_fragment"
43688. .50548
/note="assembly_fragment"
50649. .60679
/note="assembly_fragment"
60780. .68485
/note="assembly_fragment"
68586. .78448
/note="assembly_fragment"
78549. .85993
/note="assembly_fragment"
86094. .94636
/note="assembly_fragment"
94737. .114104
/note="assembly_fragment"
114205. .142862
/note="assembly_fragment"
142963. .167907
/note="assembly_fragment"
168008. .186406
/note="assembly_fragment"
clone_end:T7
vector_side:right
BASE COUNT 51735 a 38519 c 39293 g 54054 t 2805 others
ORIGIN

Query Match 9.8%; Score 20; DB 2; Length 186406;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 aagaaagctggggaagag 131
|||||
Db 154019 AAGAAAGGCTGGGGAAGAG 154000

[illegible]

```

repeat_region 11187..11411
/rpt_family="L1"
repeat_region 11412..11711
/rpt_family="Alu"
repeat_region 11694..11729
/rpt_family="A-rich"
repeat_region 11712..11972
/rpt_family="L1"
repeat_region 11973..12231
/rpt_family="Alu"
repeat_region 12232..12264
/rpt_family="L1"
repeat_region 12235..12261
/rpt_family="AT-rich"
repeat_region 12263..12610
/rpt_family="L1"
repeat_region 12611..12907
/rpt_family="Alu"
repeat_region 12908..13739
/rpt_family="L1"
repeat_region 13743..13876
/rpt_family="L1"
repeat_region 13885..13934
/rpt_family="MERL_type"
repeat_region 14077..14139
/rpt_family="MIR"
repeat_region 14576..14863
/rpt_family="Alu"
repeat_region 15079..15379
/rpt_family="Alu"
repeat_region 15648..15823
/rpt_family="Alu"
repeat_region 15853..16014
/rpt_family="Alu"
repeat_region 16014..16044
/rpt_family="(T)n"
repeat_region 16015..16319
/rpt_family="Alu"
repeat_region 16320..16454
/rpt_family="Alu"
repeat_region 17256..17384
/rpt_family="L1"
repeat_region 17389..17628
/rpt_family="Alu"
repeat_region 17639..18045
/rpt_family="ERVL"
repeat_region 18031..18465
/rpt_family="(TA)n"
repeat_region 18455..18515
/rpt_family="ERVL"
repeat_region 18516..18576
/rpt_family="Alu"
repeat_region 18643..18952
/rpt_family="Alu"
repeat_region 18923..18952
/rpt_family="(A)n"

```

Query Match 9.8%; Score 20; DB 9; Length 154037;

Best Local Similarity 100.0%; Pred. No. 2.9;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 98 agctaaacgaatttaagaaa 117
|||||
DB 9274 AGCTAAACGAATTAAAGAAA 9293

```

RESULT 8

```

AC099742 AC099742 172915 bp DNA linear HTG 20-NOV-2001
LOCUS Papio cynocephalus anubis clone RP41-167P22, WORKING DRAFT
DEFINITION SEQUENCE, 4 unordered pieces.
ACCESSION AC099742
VERSION AC099742.1 GI:17017546

```

KEYWORDS SOURCE ORGANISM

HTG; HTGS_PHASE1; HTGS_DRAFT.

olive baboon.

Papio cynocephalus anubis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Papio.

1 (bases 1 to 172915)

REFERENCE

AUTHORS

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-Q.,
Legaspi, R., Maduro, O.L., Maduro, V.B., Masiello, C., Mastrian, S.D.,
McCluskey, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y.,
Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
Tsurgou, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Zhang, L.-H. and Green, E.D.

NISC Comparative Sequencing Initiative

Unpublished

2 (bases 1 to 172915)

REFERENCE

AUTHORS

Green, E.D.

TITLE

JOURNAL

TITLE

JOURNAL

COMMENT

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc_mouse@nhgri.nih.gov

----- Project Information

Center project name: ccv

Center clone name: 167P22

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 171178 bases at least Q40

Consensus quality: 171910 bases at least Q30

Consensus quality: 172236 bases at least Q20

Insert size: 130000; agarose-fp

Quality coverage: 12.58x in Q20 bases; agarose-fp

Quality coverage: 9.47x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 2438: contig of 2438 bp in length

* 2439 2538: gap of unknown length

* 2539 8133: contig of 5595 bp in length

* 8134 8233: gap of unknown length

* 8234 40378: contig of 32145 bp in length

* 40379 40478: gap of unknown length

* 40479 172915: contig of 132437 bp in length.

FEATURES

source

1..172915

/organism="Papio cynocephalus anubis"

/db_xref="taxon:9555"

/clone="RP41-167P22"

/clone_lib="RP41"

misc_feature

1..2438

/note="assembly_fragment"

clone_end:SP6

vector_side:right

2539..8133

/note="assembly_fragment"

8234..40378

/note="assembly_fragment"

clone_end:T7

RESULT 7

AC072054 LOCUS 154037 bp DNA linear PRI 30-JAN-2002
 DEFINITION Homo sapiens BAC clone RP11-33P21 from 7, complete sequence.
 AC072054 ACCESSION
 AC072054.10 GI:18056706
 KEYWORDS HYG.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792

REFERENCE

AUTHORS Edwards, J., Du, H., Haakenson, W. and Spalding, L.
 TITLE The sequence of Homo sapiens BAC clone RP11-33P21
 JOURNAL Unpublished (2001)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (07-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (04-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (30-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On Jan 4, 2002 this sequence version replaced gi:13786481.

REFERENCE

AUTHORS Center: Washington University Genome Sequencing Center
 TITLE Center code: WUGSC
 JOURNAL Web site: <http://genome.wustl.edu/gsc>
 COMMENT Contact: sapiens@wustl.wustl.edu
 ----- Summary Statistics

 ----- Center project name: H_NH0033P21

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,

Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP13-552B11; the clone sequenced to the right is RP11-32P3, 2000 bp overlap. Actual start of this clone is at base position 22189 of RP13-552B11; actual end is at base position 2000 of RP11-32P3.

FEATURES

Location/Qualifiers
 1..154037
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7"
 /clone="RP11-33P21"
 /clone_lib="RPCI-11"
 /map="7"
 1..550
 /rpt_family="L1"
 789..1317
 /rpt_family="ERVL"
 4289..4432
 /rpt_family="MIR"
 4964..5214
 /rpt_family="Alu"
 5233..5483
 /rpt_family="L1"
 5500..5520
 /rpt_family="AT-rich"
 5994..6283
 /rpt_family="Alu"
 6337..6661
 /rpt_family="Alu"
 6636..6661
 /rpt_family="(A)n"
 6991..7114
 /rpt_family="L1"
 7354..7433
 /rpt_family="T-rich"
 7401..7634
 /rpt_family="Alu"
 7641..7759
 /rpt_family="L2"
 7797..7843
 /rpt_family="L1"
 7964..8144
 /rpt_family="MaLR"
 8148..8458
 /rpt_family="Alu"
 8461..8747
 /rpt_family="Alu"
 8755..8806
 /rpt_family="(GGGA)n"
 9513..9762
 /rpt_family="Alu"
 9768..9936
 /rpt_family="Alu"
 9938..10243
 /rpt_family="Alu"
 10278..10576
 /rpt_family="L1"
 10577..10897
 /rpt_family="Alu"
 10696..10725
 /rpt_family="AT-rich"
 10872..10897
 /rpt_family="(A)n"
 10898..11020
 /rpt_family="L1"

MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
 Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

COMMENT

Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 9, 2002 this sequence version replaced gi:17223322.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIGR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L22005

Center clone name: 627_K_17

* NOTE: This record contains 76 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1' 732: contig of 732 bp in length
 * 733 832: gap of 100 bp
 * 833 1567: contig of 735 bp in length
 * 1568 1667: gap of 100 bp
 * 1668 2347: contig of 680 bp in length
 * 2348 2447: gap of 100 bp
 * 2448 3150: contig of 703 bp in length
 * 3151 3250: gap of 100 bp
 * 3251 3958: contig of 708 bp in length
 * 3959 4058: gap of 100 bp
 * 4059 4760: contig of 702 bp in length
 * 4761 4860: gap of 100 bp
 * 4861 5573: contig of 713 bp in length
 * 5574 5673: gap of 100 bp
 * 5674 6389: contig of 716 bp in length
 * 6390 6489: gap of 100 bp
 * 6490 7173: contig of 684 bp in length
 * 7174 7273: gap of 100 bp
 * 7274 7978: contig of 705 bp in length
 * 7979 8078: gap of 100 bp
 * 8079 8798: contig of 720 bp in length
 * 8799 8898: gap of 100 bp
 * 8899 9674: contig of 776 bp in length
 * 9675 9774: gap of 100 bp
 * 9775 10478: contig of 704 bp in length
 * 10479 10578: gap of 100 bp
 * 10579 11293: contig of 715 bp in length
 * 11294 11393: gap of 100 bp
 * 11394 12101: contig of 708 bp in length
 * 12102 12201: gap of 100 bp
 * 12202 12917: contig of 716 bp in length
 * 12918 13017: gap of 100 bp
 * 13018 13746: contig of 729 bp in length
 * 13747 13846: gap of 100 bp

* 13847 14587: contig of 741 bp in length
 * 14588 14687: gap of 100 bp
 * 14688 15370: contig of 683 bp in length
 * 15371 15470: gap of 100 bp
 * 15471 16199: contig of 729 bp in length
 * 16200 16299: gap of 100 bp
 * 16300 17033: contig of 734 bp in length
 * 17034 17133: gap of 100 bp
 * 17134 17867: contig of 734 bp in length
 * 17868 17967: gap of 100 bp
 * 17968 18666: contig of 699 bp in length
 * 18667 18766: gap of 100 bp
 * 18767 19476: contig of 710 bp in length
 * 19477 19576: gap of 100 bp
 * 19577 20283: contig of 707 bp in length
 * 20284 20383: gap of 100 bp
 * 20384 21079: contig of 696 bp in length
 * 21080 21179: gap of 100 bp
 * 21180 21899: contig of 720 bp in length
 * 21900 21999: gap of 100 bp
 * 22000 22721: contig of 722 bp in length
 * 22722 22821: gap of 100 bp
 * 22822 23535: contig of 714 bp in length
 * 23536 23635: gap of 100 bp
 * 23636 24377: contig of 742 bp in length
 * 24378 24477: gap of 100 bp
 * 24478 25219: contig of 742 bp in length
 * 25220 25319: gap of 100 bp
 * 25320 26011: contig of 692 bp in length
 * 26012 26111: gap of 100 bp
 * 26112 26814: contig of 703 bp in length
 * 26815 26914: gap of 100 bp
 * 26915 27625: contig of 711 bp in length
 * 27626 27725: gap of 100 bp
 * 27726 28449: contig of 724 bp in length
 * 28450 28549: gap of 100 bp
 * 28550 29279: contig of 730 bp in length
 * 29280 29379: gap of 100 bp
 * 29380 30107: contig of 728 bp in length
 * 30108 30207: gap of 100 bp
 * 30208 30885: contig of 678 bp in length
 * 30886 30985: gap of 100 bp
 * 30986 31704: contig of 719 bp in length
 * 31705 31804: gap of 100 bp
 * 31805 32512: contig of 708 bp in length
 * 32513 32612: gap of 100 bp
 * 32613 33321: contig of 709 bp in length
 * 33322 33421: gap of 100 bp
 * 33422 34124: contig of 703 bp in length
 * 34125 34224: gap of 100 bp
 * 34225 34939: contig of 715 bp in length
 * 34940 35039: gap of 100 bp
 * 35040 35744: contig of 705 bp in length
 * 35745 35844: gap of 100 bp
 * 35845 36550: contig of 706 bp in length
 * 36551 36650: gap of 100 bp
 * 36651 37387: contig of 737 bp in length
 * 37388 37487: gap of 100 bp
 * 37488 38205: contig of 718 bp in length
 * 38206 38305: gap of 100 bp
 * 38306 39018: contig of 713 bp in length
 * 39019 39118: gap of 100 bp
 * 39119 39831: contig of 713 bp in length
 * 39832 39931: gap of 100 bp
 * 39932 40636: contig of 705 bp in length
 * 40637 40736: gap of 100 bp
 * 40737 41443: contig of 707 bp in length
 * 41444 41543: gap of 100 bp
 * 41544 42258: contig of 715 bp in length
 * 42259 42358: gap of 100 bp
 * 42359 43055: contig of 697 bp in length
 * 43056 43155: gap of 100 bp
 * 43156 43878: contig of 723 bp in length

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 13922: contig of 13922 bp in length
 * 13923 14022: gap of unknown length
 * 14023 36248: contig of 22226 bp in length
 * 36249 36348: gap of unknown length
 * 36349 66192: contig of 29844 bp in length
 * 66193 66292: gap of unknown length
 * 66293 92168: contig of 25876 bp in length
 * 92169 92268: gap of unknown length
 * 92269 122036: contig of 29768 bp in length
 * 122037 122136: gap of unknown length
 * 122137 162928: contig of 40792 bp in length.

FEATURES

source

1. 162928
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="RP43-120K11"
 /clone_lib="RP43"
 1. 13922
 /note="assembly_fragment"
 clone_end:T7
 vector_side:left"
 14023..36248
 /note="assembly_fragment"
 36349..66192
 /note="assembly_fragment"
 66293..92168
 /note="assembly_fragment"
 92269..122036
 /note="assembly_fragment"
 122137..162928
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:right"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

BASE COUNT 51528 a 29878 c 30533 g 50485 t 504 others
 ORIGIN

Query Match 14.7%; Score 30; DB 2; Length 162928;

Best Local Similarity 100.0%; Pred. No. 9.5e-06; Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ttatatacaccacaaacttggctgtctgc 35

Db 116343 TTTATACACCAACAACTTGTCTGCTC 116314

RESULT 4

G60919

LOCUS

SHGC-84306 Human Homo sapiens STS genomic, sequence tagged site.

ACCESSION

G60919

VERSION

G60919.1

KEYWORDS

STS.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 579)
 Olivier, M. and Cox, D.R.
 Unpublished
 Contact: Michael Olivier, David R. Cox
 Stanford Human Genome Center
 Stanford University School of Medicine
 4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
 Tel: (650) 320-5800
 Fax: (650) 320-5801
 Email: olivier@shgc.stanford.edu
 Primer A: TTGTGCCCCAAGAAATCTCTAA

Primer B: TCGCTTCTTTCTCCCTCCTAACT

STS size: 327

PCR Profile:

Initial incubation: 95 degrees C for 10 minutes
 Denaturation: 94 degrees C for 30 seconds
 Annealing: 60 degrees C for 30 seconds
 Polymerization: 72 degrees C for 23 seconds
 PCR Cycles: 30
 Thermal Cycler: Perkin Elmer 9700

Protocol:

Template: 25 ng
 Primer: each 1 uM
 dNTPs: each 200 uM
 AmpliTaq Gold Polymerase: 0.07 units/ul
 Total Vol: 5 ul

Buffer:

MgCl2: 2.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 8.3

BAC ends sequenced at TIGR from the RPC111 BAC library. Designed and developed at the Stanford Human Genome Center.

FEATURES source

1. 579
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="15"
 /clone_lib="Human"
 119..445
 primer_bind
 119..141
 primer_bind
 complement(423..445)
 BASE COUNT 170 a 114 c 144 g 151 t
 ORIGIN

STS

primer_bind

primer_bind

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

Qy

Db

RESULT

AC104034

LOCUS

DEFINITION

AC104034

ACCESSION

AC104034.2

VERSION

HTG: HTGS_PHASE0.

KEYWORDS

human.

SOURCE

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

COMMENT

1 (bases 1 to 61850)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Unpublished

2 (bases 1 to 61850)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhaltier, B.,

Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,

Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,

Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,

```

/note="similar to human EST W32120 (NID:g1313113)
zb97c08.r1"
misc_feature
27686. .27838
/note="similar to human EST W32085 (NID:g1313105)
zb97b08.r1"
misc_feature
complement(27733. .28143)
/note="similar to human EST N59831 (NID:g1203721)
yz77a10.s1"
misc_feature
complement(27838. .28143)
/note="similar to human EST W31561 (NID:g1312680)
zb97b08.s1"
misc_feature
complement(27887. .28143)
/note="similar to human EST N59830 (NID:g1203720)
yz77a09.s1"
misc_feature
complement(27997. .28144)
/note="similar to human EST W31628 (NID:g1312688)
zb97c08.s1"
misc_feature
28666. .28972
/note="similar to human EST R09339 (NID:g761262)
yf26d05.r1"
misc_feature
complement(29587. .29853)
/note="similar to human EST R09227 (NID:g761150)
yf26d05.s1"
misc_feature
30146. .30314
/note="similar to human EST N78042 (NID:g1240743)
yv71e06.r1"
misc_feature
complement(31170. .31350)
/note="similar to human EST N58451 (NID:g1202341)
yv71e06.s1"

repeat_region
31970. .32006
/rpt_family="L1"
repeat_region
32263. .32563
/rpt_family="ALU"
repeat_region
complement(34737. .34778)
/rpt_family="L1"
repeat_region
36558. .36709
/rpt_family="ALU"
repeat_region
complement(37630. .37740)
/rpt_family="PTR"
repeat_region
39839. .40125
/rpt_family="ALU"
repeat_region
40126. .40157
/rpt_family="L1"
repeat_region
complement(41340. .41706)
/rpt_family="ALU"
repeat_region
complement(41904. .41933)
/rpt_family="L1"
repeat_region
42348. .42409
/rpt_family="ALU"
repeat_region
complement(42942. .42982)
/rpt_family="L1"
repeat_region
complement(44380. .44423)
/rpt_family="L1"
repeat_region
44659. .44954
/rpt_family="ALU"
repeat_region
45422. .45448
/rpt_family="L1"
repeat_region
complement(45615. .45669)
/rpt_family="L1"
repeat_region
complement(45672. .45963)
/rpt_family="ALU"
repeat_region
complement(45964. .46597)
/rpt_family="L1"
repeat_region
46100. .46516
/rpt_family="L1"
repeat_region
48370. .48785
/rpt_family="L1"
repeat_region
complement(49577. .50216)
/rpt_family="L1"
repeat_region
complement(50540. .50791)
/rpt_family="THE"
repeat_region
complement(50820. .50985)
/rpt_family="THE"

```

```

repeat_region complement(51124. .51156)
/rpt_family="L1"
repeat_region 51405. .51691
/rpt_family="ALU"
repeat_region complement(51721. .51764)
/rpt_family="L1"
repeat_region complement(51765. .52058)
/rpt_family="ALU"
repeat_region complement(52252. .52467)

Query Match 14.7% Score 30; DB 9; Length 156214;
Best Local Similarity 100.0%; Pred. No. 9.5e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ttatacacacacaaactttgttcttc 35
|||||
Db 24701 TTTATACACCAACCAACTTTGTTCTGCTC 24730

RESULT 3
AC104475/c
LOCUS 162928 bp DNA linear HTG 12-DEC-2001
DEFINITION Pan troglodytes clone RP43-120K11, WORKING DRAFT SEQUENCE, 6
unordered pieces.
ACCESSION AC104475
VERSION AC104475.1 GI:17530717
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE chimpanzee
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 162928)
AUTHORS Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Hahnig, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-Q.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Masiello, C., Mastrian, S.D.,
McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Stantripop, S.,
Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L.,
Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and
Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
REFERENCE 2 (bases 1 to 162928)
AUTHORS Green, E.D.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2001) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@hgrl.nih.gov
----- Project Information
Center project name: ces
Center clone name: 120K11
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 161001 bases at least Q40
Consensus quality: 161422 bases at least Q30
Consensus quality: 161562 bases at least Q20
Insert size: 143000; agarose-fp
Insert size: 162428; sum-of-contigs
Quality coverage: 10.76x in Q20 bases; agarose-fp
Quality coverage: 9.47x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

/note="unnamed protein product"

/codon_start=1
/protein_id="CAC42678.1"
/db_xref="GI:14536714"
/translation="MESISMGMSPKSLSETCLPNGINGIKDKARKVTGVIGSGDFPKS
DLRLRGYHWVIGSRNPKASEPFFHVVDVTHEEDALTKTNILFVAIHREHYTSLW
LTHLLGKILLIDVSNMRINQYPSNAEYLASLPDLSLVKGFNVSAWALQLGPKD
ASRQVYCSNIIQARQVIELARQLNFPIDLGSSAREENLPLRLFTLWRGPVVV
AISLATFFLSFVRVDVJHPYARNOOSDFKIPIEIVNKTLPPIVAITLLSLVYLAGLL
AAAYOLYGYTRRPPWLETLQCKQLGLSLFFAMVHVAISLCPLMRRSERYLFL
NMAVQVHANISNWEVEEVRIEYISFGIMSLGLSLLAVTSIPSVSNALNWRFS
FIOSLTGYVALLISTFHVLIYGWKRAFEERYFVTPPNFVLALVLPISVILDLQLC
RYPD"

BASE COUNT 648 a 537 c 520 g 748 t

ORIGIN

Query Match 14.7% Score 30; DB 6; Length 2453;

Best Local Similarity 100.08; Pred. No. 1.6e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ttatataccacacaaacttggcttctgtc 35

Db 1631 TTTATACACCAACAACTTTGTTCTGCTC 1660

RESULT 2

HSAC002064

LOCUS

DEFINITION

AC002064

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

HSAC002064 156214 bp DNA linear PRI 09-MAY-1997
Human BAC clone RG016J04 from 7q21, complete sequence.

AC002064

AC002064.1 GI:2076723

HTG.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 156214)

Gattung,S. and Maggi,L.

The sequence of H. sapiens BAC clone RG016J04

Unpublished (1997)

2 (bases 1 to 156214)

Waterston,R.

Direct Submission

Submitted (09-MAY-1997)

Genome Sequencing Center

Department of Genetics, Washington University

St. Louis, MO 63108, USA

http://genome.wustl.edu/gsc

e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequencing of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/STB/CHR7> or send an E-mail to egreen@nhgri.nih.gov

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The

library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).
VECTOR: pBEO
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of H_RG016J04; actual end is at 156214 of H_RG016J04. The orientation of this clone is unknown.

This clone contains STS SWSS2784 (NID:g1113580) and SWSS893 (NID:g454733).

FEATURES
source

Location/Qualifiers

1..156214

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21"

/clone="RG016J04"

/clone_lib="CITB-978SK-B"

15..40

/rpt_family="LI"

326..621

/rpt_family="ALU"

complement(977..1499)

/rpt_family="LI"

3398..3421

/rpt_family="LI"

5319..5345

/rpt_family="LI"

complement(8071..8347)

/rpt_family="ALU"

complement(9406..9975)

/rpt_family="LI"

complement(10000..11285)

/rpt_family="LI"

complement(11315..11984)

/rpt_family="LI"

11666..11687

/rpt_family="LI"

complement(12010..12299)

/rpt_family="ALU"

complement(12301..13893)

/rpt_family="LI"

13897..14096

/rpt_family="LI"

complement(14777..14838)

/rpt_family="LI"

15715..15767

/note="similar to human EST T02878 (NID:g314119)"

19436..19497

/note="similar to human EST AA123941 (NID:g1682616)

mq22f09.r1"

complement(25113..25148)

/rpt_family="LI"

complement(25561..25578)

/rpt_family="LI"

complement(26832..27124)

/rpt_family="ALU"

27125..27383

/note="similar to human EST AA151807 (NID:g1720502)

2139g02.r1"

27125..27344

/note="similar to human EST AA151796 (NID:g1720491)

2139c02.r1"

complement(27178..27600)

/note="similar to human EST AA149579 (NID:g1720380)

2139c02.s1"

complement(27441..27890)

/note="similar to human EST N52554 (NID:g1193720)

yv36a11.s1"

27686..27856

misc_feature

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:55:54 ; Search time 4356.39 seconds
(without alignments)
979.943 Million cell updates/sec

Title: US-09-802-520-8

Perfect score: 204

Sequence: 1 cagagtttatacaccaccaa.....gggtcacagtaatgggatga 204

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_em.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

RESULT 1

AX155249

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

SOURCE

CDS

AX155249 Sequence

AC002064 Human BAC

AC104475 Pan trogl

G50919 SHGC-84306

AC104034 Homo sapi

AC107958 Homo sapi

AC072054 Homo sapi

AC099742 Papio cyn

AC026639 Homo sapi

AC087828 Homo sapi

AC091245 Homo sapi

AC018901 Homo sapi

AC026813 Mus muscu

AC092404 Mus muscu

G57411 SHGC-103088

AP000095 Homo sapi

AP000199 Homo sapi

Continuation (19 o

AP000239 Homo sapi

AP003894 Oryza sat

AC015494 Homo sapi

AC018988 Homo sapi

AC027334 Homo sapi

AC026670 Homo sapi

AC053543 Homo sapi

AC090668 Homo sapi

AC008401 Homo sapi

AC084373 Homo sapi

AC013281 Homo sapi

AC068015 Homo sapi

AC087478 Homo sapi

AC079075 Homo sapi

AP001705 Homo sapi

AF058089 Arabidops

AF277674 Gossypium

AJ003022 Streptomy

AL022374 Streptomy

AC005222 Homo sapi

AC102850 Mus muscu

AC003670 Homo sapi

AL138659 Arabidops

AL356014 Arabidops

AF466202 Zea mays

AC108704 Homo sapi

AC023437 Homo sapi

ALIGNMENTS

AX155249 2453 bp DNA linear PAT 22-JUN-2001

Sequence 7 from Patent WO0140276.

AX155249

AX155249.1 GI:14536713

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2453)

Afar,D.E., Hubert,R.S., Raitano,A.B., Saffran,D.C., Mitchell,S.C.,

Faris,M. and Jakobovits,A.

Serentine transmembrane antigens expressed in human prostate

cancers and uses thereof

Patent: WO 0140276-A 7 07-JUN-2001;

Urogenesys, Inc. (US)

Location/Qualifiers

1..2453

/organism="Homo sapiens"

/db_xref="taxon:9606"

355..1719

THIS PAGE BLANK (USPTO)

```

ACCESSION   AA447322
VERSION     AA447322.1  GI:3025408
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 618)
AUTHORS     Nelson,P.S., Ng,W.-L., Schummer,M., True,L.D., Liu,A.L., Bumgarner
            ,R., Ferguson,C., Dimak,A. and Hood,L.
TITLE       An expressed-sequence-tag database of the human prostate: sequence
            analysis of 1168 cDNA clones
JOURNAL     Genomics 47 (1), 12-25 (1998)
MEDLINE     98126432
COMMENT     Contact: Nelson PS
            Department of Molecular Biotechnology
            University of Washington
            Box 357730, HSB K360, 1705 NE Pacific, Seattle, WA 98195, USA
            Tel: 206 685 7336
            Fax: 206 685 7301
            Email: psnelseu.washington.edu
            Seq primer: M13 Reverse.
FEATURES    Location/Qualifiers
            source          1..618
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="SA.C08"
                        /clone_lib="PN001-Normal Human Prostate"
                        /sex="male"
                        /note="Organ: prostate; Vector: pSport1; Site_1: Not1;
                        Site_2: SalI; PolyA+ RNA was isolated from the prostate of
                        a healthy 25 year old male organ donor. 1st strand cDNA
                        was primed with oligo(dT)Not1. Double-stranded cDNA was
                        ligated to SalI adapters and directionally cloned into
                        pSport1 (Life Technologies). Average insert size is
                        1.1kb."
BASE COUNT  165 a 141 c 118 g 192 t 2 others
ORIGIN

Query Match      7.9%; Score 21; DB 9; Length 618;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1  ctcaagtctgggtatctgcaaa 21
    ||||||||||||||||||
Db  302  CTCAGTCTGGGTATCTGCAAA 282

RESULT 15
AV731102/c
LOCUS       AV731102      645 bp      mRNA      linear      EST 17-OCT-2000
DEFINITION  AV731102 HTF Homo sapiens cDNA clone HTFAYD08 5', mRNA sequence.
ACCESSION   AV731102
VERSION     AV731102.1  GI:10840523
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 645)
AUTHORS     Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
            Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
            ,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
            Chen,J., Chen,Z. and Han,Z.
TITLE       Homo sapiens cDNA HTF clones
JOURNAL     Unpublished (2000)
COMMENT     Contact: Zequang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922

```

```

Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES    Location/Qualifiers
            source          1..645
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="HTFAYD08"
                        /clone_lib="HTF"
                        /tissue_type="Hypothalamus"
                        /dev_stage="Adult"
                        /lab_host="SOLR"
                        /note="vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                        XhoI"
BASE COUNT  200 a 114 c 99 g 229 t 3 others
ORIGIN

Query Match      7.9%; Score 21; DB 9; Length 645;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1  ctcaagtctgggtatctgcaaa 21
    ||||||||||||||||||
Db  571  CTCAGTCTGGGTATCTGCAAA 551

Search completed: September 20, 2002, 08:33:54
Job time: 8718 sec

```


High quality sequence stop: 461.

FEATURES

source
1. .490
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2555116"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and cloneIDs: Soares NBHSF pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HPF-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

136 a 93 c 88 g 173 t

ORIGIN

Query Match 7.9%; Score 21; DB 9; Length 490;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctacgtctgggtatctgcaaa 21

|||||
Db 298 CTCAGTCTGGGTATCTGCAAA 278

RESULT 12

BF439596/c
LOCUS
DEFINITION nab65f07.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:3272605 3', mRNA sequence.

ACCESSION BF439596

VERSION BF439596.1 GI:11452034

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 512)

NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from Gibco

High quality sequence stop: 468.

Location/Qualifiers

1..512

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3272605"

/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Equal amounts of plasmid DNA from five normalized

libraries were mixed, and ss circles were made in vitro.

Following HAP purification, this DNA was used as tracer in

a subtractive hybridization reaction. The driver was

PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and cloneIDs: Soares NBHSF pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HPF-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 138 a 100 c 98 g 176 t

ORIGIN

Query Match 7.9%; Score 21; DB 10; Length 512;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctacgtctgggtatctgcaaa 21

|||||
Db 300 CTCAGTCTGGGTATCTGCAAA 280

RESULT 13

CNS000QB1/c
LOCUS
DEFINITION CNS000QB1 518 bp DNA linear GSS 28-JUN-1999
Arabidopsis thaliana genome survey sequence Sp6 end of BAC F11A13
of IGF library from strain Columbia of Arabidopsis thaliana,
genomic survey sequence.

ACCESSION AL085579

VERSION AL085579.1 GI:5286719

KEYWORDS GSS.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 518)

Salanoubat, M., Choisne, N., Artiguenave, F., Brottier, P., Wincker, P.,

Samson, D., Saurin, W., Weissenbach, J. and Quetier, F.

Unpublished

2 (bases 1 to 518)

Genoscope.

Direct Submission

Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

- Web : www.genoscope.cns.fr

Location/Qualifiers

1..518

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/clone_lib="IGF"

/clone="F11A13"

/note="end : Sp6"

BASE COUNT 139 a 79 c 87 g 213 t

ORIGIN

Query Match 7.9%; Score 21; DB 12; Length 518;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 aaacaagagcagaacaaagt 74

|||||
Db 230 AAACAAGAGCAGAACAAGT 210

RESULT 14

AA447322/c
LOCUS
DEFINITION AA447322 618 bp mRNA linear EST 05-APR-1998
SA.C08R PN001-Normal Human Prostate Homo sapiens cDNA clone SA.C08
5', mRNA sequence.

```

RESULT 9
AW852181
LOCUS
DEFINITION QV0-CT0225-230300-169-g06 CT0225 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW852181
VERSION AW852181.1 GI:7947698
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=&t2=QV0-CT0225-230
300-169-g06&t3=2000-03-23&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 3
High quality sequence stop: 450.
FEATURES
source
1..450
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0225"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 144 a 91 c 105 g 110 t
ORIGIN
Query Match 7.9%; Score 21; DB 9; Length 450;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctccagtctgggtatctgcaaa 21
|||||
Db 399 CTCAGTCTGGGTATCTGCAAA 419
RESULT 10
BF114610
LOCUS
DEFINITION 7J65c01.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:3391296 3', mRNA sequence.
ACCESSION BF114610
VERSION BF114610
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 455.
FEATURES
source
1..465
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145033-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HF8-9W pool 1:
738280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 129 a 91 c 81 g 164 t
ORIGIN
Query Match 7.9%; Score 21; DB 10; Length 465;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ctccagtctgggtatctgcaaa 21
|||||
Db 298 CTCAGTCTGGGTATCTGCAAA 278
RESULT 11
AW044547/c
LOCUS
DEFINITION wy83a03.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:2555116 3', mRNA sequence.
ACCESSION AW044547
VERSION AW044547.1 GI:5905076
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco

```

```

/db_xref="taxon:9606"
/clone_image="909830"
/clone_lib="NCI_CGAP_Ew1"
/tissue_type="Ewing's sarcoma"
/lab_host="DH10B"
/note=vector: PAMP10; mRNA made from Ewing's sarcoma,
cDNA made by oligo-dT priming. Non-directionally cloned.
Size selected on agarose gel, average insert size 600 bp.
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT      67 a   63 c   61 g   92 t
ORIGIN

Query Match
Best Local Similarity  7.9%; Score 21; DB 9; Length 283;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1  ctacgtctgggtatctgcaaa 21
      |||||||
Db  90  CTCAGTCTGGGTATCTGCAAA 70

RESULT 7
AW852172
LOCUS      AW852172          407 bp      mRNA      linear      EST 19-MAY-2000
DEFINITION QVO-CT0225-230300-169-b05 CT0225 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW852172
VERSION    AW852172.1  GI:7947689
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 407)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
            Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact:  Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email:  asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=6t2-QV0-CT0225-230
300-169-b05&t3=2000-03-23&t4=1)
Seq primer:  puc 18 forward
High quality sequence stop: 12
High quality sequence stop: 407.
Location/Qualifiers
            1..407
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="CT0225"
            /dev_stage="Adult"
            /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
            SmaI; A mini-library was made by cloning products derived
            from ORESTES PCR (U.S. Letters Patent application No. 196
            ,716 - Ludwig Institute for Cancer Research) profiles
            into the puc 18 vector. Reverse transcription of tissue
            mRNA and cDNA amplification were performed under low
            stringency conditions."

FEATURES
            source
            1..407
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="CT0225"
            /dev_stage="Adult"
            /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
            SmaI; A mini-library was made by cloning products derived
            from ORESTES PCR (U.S. Letters Patent application No. 196
            ,716 - Ludwig Institute for Cancer Research) profiles
            into the puc 18 vector. Reverse transcription of tissue
            mRNA and cDNA amplification were performed under low
            stringency conditions."

```

```

BASE COUNT      127 a   86 c   96 g   98 t
ORIGIN

Query Match
Best Local Similarity  7.9%; Score 21; DB 9; Length 407;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1  ctacgtctgggtatctgcaaa 21
      |||||||
Db  375  CTCAGTCTGGGTATCTGCAAA 395

RESULT 8
BF372093/c
LOCUS      BF372093          407 bp      mRNA      linear      EST 24-NOV-2000
DEFINITION QV2-FT0010-090800-303-g05 FT0010 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF372093
VERSION    BF372093.1  GI:11334118
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 407)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
            Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact:  Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email:  asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=QV2&t2-QV2-FT0010-
090800-303-g05&t3=2000-08-09&t4=1)
Seq primer:  puc 18 forward
High quality sequence stop: 407.
Location/Qualifiers
            1..407
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="FT0010"
            /dev_stage="Adult"
            /note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
            Site_2: SmaI; A mini-library was made by cloning products
            derived from ORESTES PCR (U.S. Letters Patent application
            No. 196,716 - Ludwig Institute for Cancer Research)
            profiles into the puc 18 vector. Reverse transcription of
            tissue mRNA and cDNA amplification were performed under
            low stringency conditions."
            118 a   75 c   64 g   150 t

BASE COUNT      118 a   75 c   64 g   150 t
ORIGIN

Query Match
Best Local Similarity  7.9%; Score 21; DB 10; Length 407;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1  ctacgtctgggtatctgcaaa 21
      |||||||
Db  341  CTCAGTCTGGGTATCTGCAAA 321

```

```

MEDLINE
REFERENCE
AUTHORS
20584722
2 (bases 1 to 978)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEMS Lett. 487 (1), 3-12 (2000)
20584711
3 (bases 1 to 978)
Direct Submission
Genoscope.
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
source
Location/Qualifiers
1..978
/organism="Kluyveromyces marxianus"
/strain="CBS 712"
/variety="marxianus"
/db_xref="taxon:4911"
/clone="XAZ0RAA001H01"
/clone_lib="XAZ0AA"
/note="end : T7"
BASE COUNT 210 a 189 c 212 g 365 t 2 others
ORIGIN
1..978
/organism="Kluyveromyces marxianus"
/strain="CBS 712"
/variety="marxianus"
/db_xref="taxon:4911"
/clone="XAZ0RAA001H01"
/clone_lib="XAZ0AA"
/note="end : T7"

Query Match 8.38; Score 22; DB 12; Length 978;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 gcaaaacaaagagcaagaacaaa 72
|||||
Db 379 GCATAACAGAGCAGACACAAA 358

RESULT 5
LOCUS BF360898 248 bp mRNA linear EST 22-NOV-2000
DEFINITION QV4-OT0067-060700-277-f06 OT0067 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF360898
VERSION BF360898.1 GI:11319970
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 248)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE
20202663
COMMENT Contact: Simpson A.J.G.

MEDLINE
REFERENCE
AUTHORS
20584722
2 (bases 1 to 978)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEMS Lett. 487 (1), 3-12 (2000)
20584711
3 (bases 1 to 978)
Direct Submission
Genoscope.
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
source
Location/Qualifiers
1..978
/organism="Kluyveromyces marxianus"
/strain="CBS 712"
/variety="marxianus"
/db_xref="taxon:4911"
/clone="XAZ0RAA001H01"
/clone_lib="XAZ0AA"
/note="end : T7"

Query Match 7.9%; Score 21; DB 10; Length 248;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctcaagtcgggtatctgcacaa 21
|||||
Db 66 CTCAGTCGTGGGTATCTGCAAA 86

RESULT 6
LOCUS AA491988/c 283 bp mRNA linear EST 19-AUG-1997
DEFINITION ne72e08.s1 NCI_CGAP_Ew1 Homo sapiens cDNA clone IMAGE:909830, mRNA
sequence.
ACCESSION AA491988
VERSION AA491988.1 GI:2221550
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 283)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 880 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 276.
Location/Qualifiers
1..283
/organism="Homo sapiens"

MEDLINE
REFERENCE
AUTHORS
20584722
2 (bases 1 to 978)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEMS Lett. 487 (1), 3-12 (2000)
20584711
3 (bases 1 to 978)
Direct Submission
Genoscope.
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
source
Location/Qualifiers
1..248
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="OT0067"
/dev_stage="Adult"
/note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 79 a 42 c 54 g 73 t
ORIGIN
1..248
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="OT0067"
/dev_stage="Adult"
/note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

```

```

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl-QV4st2-QV4-OT0067-
060700-277-f06&t3=2000-07-06&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 248.
Location/Qualifiers
1..248
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="OT0067"
/dev_stage="Adult"
/note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 79 a 42 c 54 g 73 t
ORIGIN
1..248
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="OT0067"
/dev_stage="Adult"
/note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

```

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 132 a 84 c 61 g 116 t
ORIGIN

Query Match 32.1%; Score 85; DB 9; Length 393;
Best Local Similarity 100.0%; Pred. No. 5.1e-33;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 ctgaataaaactgaattctctccagtttaaacgattgtcactgaaggatagaagtgcac 240
|||||
Db 267 CTGAATAAACTGAATTCTCTCCAGTTAAAGCATTTGCTCACTGAAGGATAGAAGTGAC 326
|||||

Qy 241 tgcaggaggagaaagtaagccaagg 265
|||||
Db 327 TGCCAGGAGGAGGAAGTAAGCCCAAGG 351
|||||

RESULT 2
BM431438/c
LOCUS BM431438 558 bp mRNA linear EST 31-JAN-2002
DEFINITION 1Duo16F03 Bos taurus Duodenum #1 library Bos taurus cDNA, mRNA
sequence.
ACCESSION BM431438
VERSION BM431438.1 GI:18453160
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
AUTHORS Hansen, C., Fu, A., Meng, Y., Li, C., Okine, E., Senses, C.W., Gordon
P.M.K. and Moore, S.S.
TITLE Gene Expression Profiling of the Bovine Gastrointestinal Tract
JOURNAL Unpublished (2002)
COMMENT Contact: Dr. Stephen Moore
Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca
Insert Length: 558 Std Error: 0.00
POLYA-No. Location/Qualifiers

FEATURES source
1..558
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bos Taurus Duodenum #1 library"
/tissue_type="Smooth muscle"
/cell_type="Simple columnar epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMRF/strain"
/note="Organ: Intestine/duodenum; Vector: Uni-22APXR;
Site_1: EcoRI; Site_2: Xho I"
BASE COUNT 152 a 109 c 108 g 189 t
ORIGIN

Query Match 21.9%; Score 58; DB 10; Length 558;
Best Local Similarity 100.0%; Pred. No. 4.2e-19;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 33. ccagaattacaattgaggcgaacaaagagcaagaacaaagtgtgtgtgtataaaa 90
|||||

Db 199 CCAGAATTACAAATTGAGGCGCAACAAAGCAAGCAAGAAAGTTTGGTGGTGATATAAA 142
|||||

..

RESULT 3
BI359597/c
LOCUS BI359597 595 bp mRNA linear EST 01-AUG-2001
DEFINITION 384188 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BI359597
VERSION BI359597.1 GI:15055625
KEYWORDS EST.
SOURCE pig.

ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
AUTHORS 1 (bases 1 to 595)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, W.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
and Keele, J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGAGC
Plate: 132 row: H column: 3
Seq primer: ATTAGGTGACACTATAG.

FEATURES source
1..595
Location/Qualifiers
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 143 a 138 c 126 g 188 t
ORIGIN

Query Match 11.3%; Score 30; DB 10; Length 595;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 236 gtgactgcaggagggaagtaagccaagg 265
|||||

Db 593 GTGACTGCCAGGAGGAAAGTAAGCCAAGG 564
|||||

RESULT 4
CNS071AX/c
LOCUS CNS071AX 978 bp DNA linear GSS 06-JUL-2001
DEFINITION T7 end of clone XAZ0AA001H01 of library XAZ0AA from strain CBS 712
of Kluyveromyces marxianus, genomic survey sequence.

ACCESSION AL424587
VERSION AL424587.1 GI:12207881
KEYWORDS GSS.
SOURCE Kluyveromyces marxianus.

ORGANISM Kluyveromyces marxianus
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
AUTHORS 1 (bases 1 to 978)
Llorente, B., Maipertuy, A., Blandin, G., Artiguenave, F., Wincker, P.
and Dujon, B.

TITLE Genomic exploration of the hemiascomycetous yeasts: 12.
JOURNAL Kluyveromyces marxianus var. marxianus
FEBS Lett. 487 (1), 71-75 (2000)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 08:33:51 ; Search time 5053.42 seconds
(without alignments)
707.777 Million cell updates/sec

Title: US-09-802-520-7
Perfect score: 265
Sequence: 1 ctcaagtctgtgtatctgca.....ggagggaagtaagccaagg 265

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85	32.1	393	AI016724	AI016724 ov03d03.x
2	58	21.9	558	BM431438	BM431438 IDv016F03
3	30	11.3	595	BI359597	BI359597 384188 MA
4	22	8.3	978	AL424687	AL424687 T7 end of
5	21	7.9	248	BF360898	BF360898 QV4-OT006
6	21	7.9	283	AA491988	AA491988 ne72e08.s
7	21	7.9	407	AW852172	AW852172 QV0-CT022
8	21	7.9	407	BF372093	BF372093 QV2-FT001
9	21	7.9	450	AW852181	AW852181 QV0-CT022
10	21	7.9	465	BF114610	BF114610 7J6Sc01.x
11	21	7.9	490	AW044547	AW044547 wy83a03.x
12	21	7.9	512	BF439596	BF439596 nab65f07
13	21	7.9	518	CNS000B1	AL085579 Arabidops
14	21	7.9	618	AA447322	AA447322 SA.C08R.P
15	21	7.9	645	AV731102	AV731102 AV731102
16	21	7.9	666	AV731549	AV731549 AV731549
17	21	7.9	802	AW971939	AW971939 EST384148

c 18	20	7.5	198	12	AQ544394
c 19	20	7.5	278	9	BB479717
c 20	20	7.5	322	9	AW011108
c 21	20	7.5	488	12	AQ833520
c 22	20	7.5	873	10	BG403540
c 23	19	7.2	130	10	R26307
c 24	19	7.2	163	10	BF661645
c 25	19	7.2	178	10	BF452991
c 26	19	7.2	180	9	BB213248
c 27	19	7.2	199	9	AV232888
c 28	19	7.2	203	10	BE687167
c 29	19	7.2	227	9	AV020448
c 30	19	7.2	229	9	AV027007
c 31	19	7.2	230	12	BH053900
c 32	19	7.2	233	9	AV273266
c 33	19	7.2	235	9	BB056822
c 34	19	7.2	235	9	BB296994
c 35	19	7.2	236	9	BB048145
c 36	19	7.2	238	9	AV021787
c 37	19	7.2	240	9	AV353049
c 38	19	7.2	244	9	AV016593
c 39	19	7.2	244	9	BB059246
c 40	19	7.2	246	9	AV314454
c 41	19	7.2	248	9	BB296861
c 42	19	7.2	251	9	AV013960
c 43	19	7.2	253	9	AV026902
c 44	19	7.2	253	9	AV228977
c 45	19	7.2	258	10	BE848896

ALIGNMENTS

RESULT 1

AI016724
LOCUS
DEFINITION
ov03d03.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1636229 3', mRNA linear EST 27-AUG-1998
mRNA sequence.
AI016724
ACCESSION
AI016724.1 GI:3231060
VERSION
AI016724.1
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 393)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
CDNA Library Arrayed by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 305 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES
source

1..393
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1636229"
/clone_lib="NCI_CGAP_Kid3"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pTT73D-Pac (Pharmacia) with a modified polylinker; Site:1: Not 1; Site:2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors

THIS PAGE BLANK (USPTO)

Query Match 6.0%; Score 16; DB 2; Length 1443;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 aattaaaaacatggaaa 145
|||||
Db 686 AATTAAACATGGAAA 701

Search completed: September 20, 2002, 09:51:17
Job time: 11865 sec

;; APPLICANT: Guegler, Karl J.
;; TITLE: Hawkins, Phillip R.
;; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES
;; NUMBER OF SEQUENCES: 9
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Incyte Pharmaceuticals, Inc.
;; STREET: 3174 Porter Drive
;; CITY: Palo Alto
;; STATE: CA
;; COUNTRY: U.S.
;; ZIP: 94304
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq Version 1.5
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/541,228
;; FILING DATE:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/712,709
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Billings, Lucy J.
;; REGISTRATION NUMBER: 36,749
;; REFERENCE/DOCKET NUMBER: PF-0118 US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-855-0555
;; TELEFAX: 415-845-4166
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2161 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; IMMEDIATE SOURCE:
;; LIBRARY:
;; CLONE: Consensus
US-09-541-228-4

Query Match 6.4%; Score 17; DB 4; Length 2161;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 144 aagttacttatgagcaga 160
|||||
Db 2027 AAGTACTTATGAGCAGA 2011

RESULT 14
US-09-328-111-84/c
;; Sequence 84, Application US/09328111
;; Patent No. 6262333
;; GENERAL INFORMATION:
;; APPLICANT: Endege, Wilson O.
;; APPLICANT: Steinmann, Kathleen E.
;; APPLICANT: Astle, Jon H.
;; APPLICANT: Burgess, Christopher C.
;; APPLICANT: Bushnell, Steven E.
;; APPLICANT: Carroll III, Eddie
;; APPLICANT: Catino, Theodore J.
;; APPLICANT: Dertl, Adnan
;; APPLICANT: Ford, Donna M.
;; APPLICANT: Lewis, Marcia E.
;; APPLICANT: Monahan, John E.
;; APPLICANT: Schlegel, Robert
;; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;; TITLE OF INVENTION: PRODUCTS
;; FILE REFERENCE: CCD-257 (US)
;; CURRENT APPLICATION NUMBER: US/09/328,111
;; CURRENT FILING DATE: 1999-06-08

;; EARLIER APPLICATION NUMBER: US 60/088,801
;; EARLIER FILING DATE: 1998-06-10
;; NUMBER OF SEQ ID NOS: 850
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 84
;; LENGTH: 558
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-328-111-84

Query Match 6.0%; Score 16; DB 4; Length 558;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 249 gggaaagtaagcgaag 264
|||||
Db 125 GGGAAAGTAAGCCAAG 110

RESULT 15
US-08-403-853-19
;; Sequence 19, Application US/08403853
;; Patent No. 5844094
;; GENERAL INFORMATION:
;; APPLICANT: HUDSON, Peter J.
;; APPLICANT: LAH, Maria
;; APPLICANT: KORRT, Alex A.
;; APPLICANT: IRVING, Robert A.
;; APPLICANT: ATWELL, John L.
;; APPLICANT: MALBY, Robyn L.
;; APPLICANT: POWER, Barbara E.
;; APPLICANT: COLMAN, Peter M.
;; TITLE OF INVENTION: TARGET BINDING POLYPEPTIDE
;; NUMBER OF SEQUENCES: 25
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Foley & Lardner
;; STREET: 3000 K Street, N.W., Suite 500
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20007-5109
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/403,853
;; FILING DATE: 30-MAY-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: WO PCT/AU93/00491
;; FILING DATE: 24-SEP-1993
;; APPLICATION NUMBER: AU PL 4973
;; FILING DATE: 25-SEP-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BENT, Stephen A.
;; REGISTRATION NUMBER: 29,768
;; REFERENCE/DOCKET NUMBER: 16786/189/CHAC
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)672-5300
;; TELEFAX: (202)672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 19:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1443 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-08-403-853-19

NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 00786/327001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1975 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 127...1404
US-09-185-370-1

Query Match 6.4%; Score 17; DB 3; Length 1975;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 144 aagtacttatgagcaga 160
|||||
Db 1914 AAGTACTTATGAGCAGA 1898

RESULT 11
US-08-712-709-4/c
Sequence 4, Application US/08712709
Patent No. 5863780
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Guegler, Karl J.
APPLICANT: Hawkins, Phillip R.
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/712,709
FILING DATE: Filed Herewith
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0118 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2161 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
LIBRARY: Consensus
CLONE: Consensus
US-08-712-709-4

Query Match 6.4%; Score 17; DB 2; Length 2161;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 144 aagtacttatgagcaga 160
|||||
Db 2027 AAGTACTTATGAGCAGA 2011

RESULT 12
US-09-111-444-4/c
Sequence 4, Application US/09111444
Patent No. 6045792
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Guegler, Karl J.
APPLICANT: Hawkins, Phillip R.
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/111,444
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/712,709
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0118 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2161 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
IMMEDIATE SOURCE:
LIBRARY: Consensus
CLONE: Consensus
US-09-111-444-4

Query Match 6.4%; Score 17; DB 3; Length 2161;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 144 aagtacttatgagcaga 160
|||||
Db 2027 AAGTACTTATGAGCAGA 2011

RESULT 13
US-09-541-228-4/c
Sequence 4, Application US/09541228
Patent No. 6232077
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice

; SEQ ID NO 10
; LENGTH: 180
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-323-873A-10

Query Match 7.9%; Score 21; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 0.054;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctcagctcgggtatctgcaaa 21
|||||
Db 29 CTCAGCTCTGGGTATCTGCAAA 9

RESULT 8
US-09-439-313-288
; Sequence 288, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqui
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 288
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-439-313-288

Query Match 7.9%; Score 21; DB 4; Length 301;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctcagctcgggtatctgcaaa 21
|||||
Db 91.ctcagctcgggtatctgcaaa 111

RESULT 9
US-08-852-743-1/c
; Sequence 1, Application US/08852743
; Patent No. 5830699
; GENERAL INFORMATION:
; APPLICANT: Force, Thomas
; APPLICANT: Kyriakis, John M.
; APPLICANT: Pombo, Celia M.
; APPLICANT: Bonventre, Joseph
; TITLE OF INVENTION: SOK-1 AND METHODS OF USE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,743
; FILING DATE: 7-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/016,774
; FILING DATE: 7-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00786/327001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 127...1404
US-08-852-743-1

Query Match 6.4%; Score 17; DB 2; Length 1975;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 144 aagtacttatgagcaga 160
|||||
Db 1914 AAGTACTTATGAGCAGA 1898

RESULT 10
US-09-185-370-1/c
; Sequence 1, Application US/09185370
; Patent No. 6093560
; GENERAL INFORMATION:
; APPLICANT: Force, Thomas
; APPLICANT: Kyriakis, John M.
; APPLICANT: Pombo, Celia M.
; APPLICANT: Bonventre, Joseph
; TITLE OF INVENTION: SOK-1 AND METHODS OF USE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/185,370
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/852,743
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(250)
; OTHER INFORMATION: n = A,T,C or G
US-09-439-313-213

Query Match 35.1%; Score 93; DB 4; Length 250;
Best Local Similarity 100.0%; Pred. No. 9.5e-40;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 173 agttagactgaataaaactgaattctctccagtttaagcattgctcactgaaggata 232

Db 26 agtgtagactgaataaaactgaattctctccagtttaagcattgctcactgaaggata 85

Qy 233 gaagtactgcagaggaggaagtaagccaagg 265

Db 86 gaagtactgcagaggaggaagtaagccaagg 118

RESULT 5

US-09-030-607-215
; Sequence 215, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 215:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 366 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-030-607-215

Query Match 32.5%; Score 86; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 4.5e-36;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 180 actgaataaaactgaattctctccagtttaagcattgctcactgaaggatagaagtga 239

Db 33 ACTGAATAAACTGAATTCTCTCCAGTTTAAAGCATTGCTCAGTGAAGGATAGAAGTGA 92

Qy 240 ctgccaggagggaagtaagccaagg 265

Db 93 CTGCCAGGAGGAGAAAGTAAGCCAAGG 118

RESULT 6

US-09-439-313-215
; Sequence 215, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqui
; APPLICANT: Read, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9

; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 366
; TYPE: DNA
; ORGANISM: Homo sapien

; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(366)
; OTHER INFORMATION: n = A,T,C or G
US-09-439-313-215

Query Match 32.5%; Score 86; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 4.5e-36;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 180 actgaataaaactgaattctctccagtttaagcattgctcactgaaggatagaagtga 239

Db 33 actgaataaaactgaattctctccagtttaagcattgctcactgaaggatagaagtga 92

Qy 240 ctgccaggagggaagtaagccaagg 265

Db 93 ctgccaggagggaagtaagccaagg 118

RESULT 7

US-09-323-873A-10/c
; Sequence 10, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:

; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Safran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; FILE REFERENCE: 129.160S02
; CURRENT APPLICATION NUMBER: US/09/323,873A
; CURRENT FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0

Db 487 CTGCAAAAGATCCGAATTTACAATTTGAGGCAAAACAGCAAGCAAAAGATTTGGTGG 428
Qy 82 tgtataaaatctgtagtactctctcaaaagctcgtttccatccataaaattaaaaacatg 141
Db 427 TGTATAAAATCTGTAGTACTTCTCTCAAAAGCTGTTCCATCCATAAATTAACATG 368
Qy 142 gaaagtacttatgacagacagacatatccaaagtgtagactgaataaaactgaattctct 201
Db 367 GAAAGTACTTATGACGACAGCAGACATATCCAAAGTGTAGACTGAATAAAACTGAATTTCT 308
Qy 202 ccagtttaaacattctcactgaaggatagaagtgaactgactgcagagggaagaaagcc 261
Db 307 CCAGTTTAAAGCATTTGCTACATGAAGGATAGTAGTACTGCCAGAGGGAAGTAAGCC 248
Qy 262 aagg 265
Db 247 AAGG 244

RESULT 2
US-09-323-873A-7/c
; Sequence 7, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; FILE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF
; FILE REFERENCE: 129.160S02
; CURRENT APPLICATION NUMBER: US/09/323,873A
; CURRENT FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-323-873A-7

Query Match 58.5%; Score 155; DB 4; Length 521;
Best Local Similarity 100.0%; Pred. No. 3e-72;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 111 aagctctttccatcataaaattaaaacatgaagtacttatgacagacgacatc 170
Db 521 AAGCTCGTTTCCATCCATAAATTAACATGAAGAGTACTTATGACGAGGCACATATC 462
Qy 171 caagtgtagactgaataaaactgaattctctccagtttaaaagcattgtcactgaaggga 230
Db 461 CAAGTGTAGACTGAATAAACTGAATTTCTCTCCAGTTTAAAGCATTTGCTCACTGAAGGA 402
Qy 231 tagaagtgaactccagagggaagaaagtaagccaagg 265
Db 401 TAGAAGTGACTCCAGAGGGAAGTAAGTAAAGCAAGG 367

RESULT 3
US-09-030-607-213
; Sequence 213, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.

; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 213:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 250 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-030-607-213

Query Match 35.1%; Score 93; DB 4; Length 250;
Best Local Similarity 100.0%; Pred. No. 9.5e-40;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 173 agttagactgaataaaactgaattctctccagtttaaaagcattgtcactgaaggata 232
Db 26 AGTGTAGACTGAATAAACTGAATTTCTCCAGTTTAAAGCATTTGCTCACTGAAGGATA 85
Qy 233 gaagtgaactgccaggaggaaagtaagccaagg 265
Db 86 GAAGTGACTGCCAGGAGGGAAGTAAGCCAAGG 118

RESULT 4
US-09-439-313-213
; Sequence 213, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqul
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 213
; LENGTH: 250

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:51:16 ; Search time 139.75 Seconds
(without alignments)
465.781 Million cell updates/sec

Title: US-09-802-520-7
Perfect score: 265
Sequence: 1 ctccagtcgtgggtatctgca.....ggagggaagtaagccaagg 265

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCUTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	244	92.1	1213	3	US-09-083-521-3
C 2	155	58.5	521	4	US-09-323-873A-7
C 3	93	35.1	250	4	US-09-030-607-213
C 4	93	35.1	250	4	US-09-439-313-213
C 5	86	32.5	366	4	US-09-030-607-215
C 6	86	32.5	366	4	US-09-439-313-215
C 7	21	7.9	180	4	US-09-323-873A-10
C 8	21	7.9	301	4	US-09-439-313-288
C 9	17	6.4	1975	3	US-08-852-743-1
C 10	17	6.4	1975	3	US-09-185-370-1
C 11	17	6.4	2161	2	US-08-712-709-4
C 12	17	6.4	2161	3	US-09-111-444-4
C 13	17	6.4	2161	4	US-08-541-228-4
C 14	16	6.0	558	4	US-09-328-111-84
C 15	16	6.0	1443	2	US-08-403-853-19
C 16	16	6.0	2917	2	US-08-624-581-3
C 17	16	6.0	4299	1	US-08-264-002-1
C 18	15	5.7	54	1	US-08-270-805C-20
C 19	15	5.7	54	2	US-08-410-654B-20
C 20	15	5.7	54	2	US-08-474-851-20
C 21	15	5.7	54	2	US-08-481-560-20
C 22	15	5.7	54	4	US-08-643-810A-20
C 23	15	5.7	60	1	US-08-270-805C-19
C 24	15	5.7	60	2	US-08-410-654B-19
C 25	15	5.7	60	2	US-08-474-851-19
C 26	15	5.7	60	2	US-08-481-560-19
C 27	15	5.7	60	3	US-08-170-113-11

28	15	5.7	60	4	US-08-643-810A-19	Sequence 19, Appl
29	15	5.7	60	4	US-09-552-613-11	Sequence 11, Appl
C 30	15	5.7	496	4	US-09-123-912-103	Sequence 103, App
C 31	15	5.7	511	1	US-07-903-466-45	Sequence 45, Appl
C 32	15	5.7	511	5	PCT-US93-05794-45	Sequence 45, Appl
C 33	15	5.7	572	4	US-09-328-111-206	Sequence 206, App
C 34	15	5.7	586	4	US-09-385-982-139	Sequence 139, App
C 35	15	5.7	632	4	US-09-328-111-510	Sequence 510, App
C 36	15	5.7	774	4	US-08-858-207A-177	Sequence 177, App
C 37	15	5.7	949	3	US-08-714-918-12	Sequence 12, Appl
C 38	15	5.7	949	4	US-09-265-315-12	Sequence 12, Appl
C 39	15	5.7	949	4	US-09-266-417-12	Sequence 12, Appl
C 40	15	5.7	985	3	US-08-961-083-7	Sequence 7, Appl
C 41	15	5.7	1682	1	US-08-503-133A-8	Sequence 8, Appl
C 42	15	5.7	1682	2	US-08-576-775A-8	Sequence 8, Appl
C 43	15	5.7	1682	2	US-08-972-498-8	Sequence 8, Appl
C 44	15	5.7	1682	3	US-08-899-545-8	Sequence 8, Appl
C 45	15	5.7	1682	3	US-08-899-545-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-09-083-521-3/c
; Sequence 3, Application US/09083521
; Patent No. 6048970
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,521
; FILING DATE: Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0527 US
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1213 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT10
; CLONE: 1691243
US-09-083-521-3

Query Match 92.1% Score 244; DB 3; Length 1213;
Best Local Similarity 100.0%; Pred. No. 6.5e-119;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 22 ctgcaaaagatccagaattacgaaggcgaagcaagcaagcaagtttggtg 81

PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
XX 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
XX WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 50; Page 173; 263pp; English.
XX
XX The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (ptp). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AA082000 to AA082020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 250 BP; 81 A; 53 C; 42 G; 69 T; 5 other;

Query Match 35.1%; Score 93; DB 21; Length 250;
Best Local Similarity 100.0%; Pred. No. 4.3e-37;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 173 adgtagactgaataaaactgaattctctccagtttaagaagcattgctcactgaaggata 232
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
26 adgtagactgaataaaactgaattctctccagtttaagaagcattgctcactgaaggata 85
QY 233 gaagtgaactgccaggagggaagaaagtaagccaagg 265
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
86 gaagtgaactgccaggagggaagaaagtaagccaagg 118

Search completed: September 20, 2002, 09:59:56
Job time: 11714 sec

```
Oy 93 tgtagtactctctcctcaaaagctggtttccatccataaataaacatggaagtacttta 152
Db 1502 TGTAGTACTCTCTCTCAAAAGCTCGTTTCCATCCATAAATTAACATGGAAGTACTT 1443
Oy 153 tgacagagcagacatatccaaagttaga 180
Db 1442 TGACGAGAGCGACATATCCCAAGTGTAGA 1415

RESULT 13
AAV61347
ID AAV61347 standard; cDNA; 250 BP.
XX AC
XX AAV61347;
XX DT
XX 06-JAN-1999 (first entry)
XX DE Reverse DNA sequence of prostate tumour clone 8-b5.
XX KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
XX OS Homo sapiens.
XX PN WO9837093-A2.
XX PD 27-AUG-1998.
XX PF 25-FEB-1998; 98WO-US03492.
XX PR 09-FEB-1998; 98US-0020956.
XX PR 25-FEB-1997; 97US-0806099.
XX PR 01-AUG-1997; 97US-0904804.
XX PA (CORI-) CORIXA CORP.
XX PI Dillon DC, Xu J;
XX PI WPI; 1998-609886/51.
XX PT Polypeptides comprising immunogenic portions of prostate proteins -
XX PT used in a vaccine for the treatment of prostate cancer
XX PS Claim 12; Page 121; 130pp; English.
XX CC The present sequence is a DNA which encodes an immunogenic portion
XX CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
XX CC can be used as a vaccine for the treatment of prostate cancer. The DNA
XX CC was identified by analysis of a subtracted cDNA library obtained by
XX CC subtracting a prostate tumour cDNA expression library with a normal
XX CC tissue cDNA library.
XX SQ Sequence 250 BP; 81 A; 53 C; 42 G; 69 T; 5 other;

Query Match 35.1%; Score 93; DB 19; Length 250;
Best Local Similarity 100.0%; Pred. No. 4.3e-37;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 173 agttagactgaataaaactgaattctctccagtttaagcattgctcactgaaggata 232
Db 26 agttagactgaataaaactgaattctctccagtttaagcattgctcactgaaggata 85

Oy 233 gaagtactgccaggagggaagtaagccaagg 265
Db 86 gaagtactgccaggagggaagtaagccaagg 118

RESULT 14
AAV58683
ID AAV58683 standard; cDNA; 250 BP.
XX AC
XX AAV58683;
XX DT
XX 13-JUN-2000 (first entry)
XX DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:213.
XX KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
XX KW immunogenic; cytostatic; vaccine; ss.
XX OS Homo sapiens.
XX PN WO200004149-A2.
XX PN
```

```
DT 08-DEC-1998 (first entry)
XX Prostate tumour specific gene clone 8-b5.
XX Prostate tumour specific gene; human; prostate cancer; detection;
XX KW therapy; ss.
XX OS Homo sapiens.
XX PN WO9837418-A2.
XX PD 27-AUG-1998.
XX PF 25-FEB-1998; 98WO-US03690.
XX PR 09-FEB-1998; 98US-0904809.
XX PR 25-FEB-1997; 97US-0806596.
XX PR 01-AUG-1997; 97US-0904809.
XX PA (CORI-) CORIXA CORP.
XX PI Dillon DC, Xu J;
XX PI WPI; 1998-480805/41.
XX PT Novel human prostate specific tumour protein and fragments - useful
XX PT for detecting and treating prostate cancers
XX PS Claim 1; Page 130; 141pp; English.
XX CC This sequence represents a human prostate tumour specific gene, and can
XX CC be used in the method of the invention. The method is for detecting
XX CC prostate cancer comprises contacting a biological sample with an agent
XX CC able to bind an immunogenic portion of a prostate protein (such as
XX CC encoded by this sequence). An antibody which binds to an immunogenic
XX CC portion of the prostate protein, and the method can be used to detect,
XX CC monitor progression of, or treat prostate cancers. The antibody may
XX CC also be conjugated to a therapeutic agent for use in therapy of prostate
XX CC cancers.
XX SQ Sequence 250 BP; 81 A; 53 C; 42 G; 69 T; 5 other;

Query Match 35.1%; Score 93; DB 19; Length 250;
Best Local Similarity 100.0%; Pred. No. 4.3e-37;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 173 agttagactgaataaaactgaattctctccagtttaagcattgctcactgaaggata 232
Db 26 agttagactgaataaaactgaattctctccagtttaagcattgctcactgaaggata 85

Oy 233 gaagtactgccaggagggaagtaagccaagg 265
Db 86 gaagtactgccaggagggaagtaagccaagg 118

RESULT 15
AAO6446
ID AAO6446 standard; cDNA; 250 BP.
XX AC
XX AAO6446;
XX DT
XX 13-JUN-2000 (first entry)
XX DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:213.
XX KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
XX KW immunogenic; cytostatic; vaccine; ss.
XX OS Homo sapiens.
XX PN WO200004149-A2.
XX PN
```



```
XX WO200172962-A2.
PN
XX
XX
XX PD 04-OCT-2001.
XX
XX PF 23-MAR-2001; 2001WO-US09410.
XX
XX PR 24-MAR-2000; 2000US-191929P.
XX
XX PA (SAAT/) SAATCIOGLU F.
XX
XX PI Saaticloglu F;
XX
XX DR WPI; 2001-662926/76.
XX
XX New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids
PT
XX Claim 6; Fig 4F; 114pp; English.
XX
XX The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence represents exon 6 of a prostate specific protein,
CC Six-Transmembrane Protein of Prostate 1, STMP1, alternatively
CC spliced version, ORP2.
XX
XX Sequence 148 BP; 36 A; 29 C; 26 G; 57 T; 0 other;
XX
XX Query Match 55.8%; Score 148; DB 22; Length 148;
XX Best Local Similarity 100.0%; Pred. No. 8.9e-65;
XX Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 33 ccgaattacaattgagggcaaaacaaagagcaagaacaaagtgtgtgtataaaatc 92
DB 148 CCAGAATTACAAATTGAGGCGCAAAACAAAGAGCAAGAACAAGTTTGGTGTGTATAAAATC 89
QY 93 ttagtactcttcctcaaaagctgtttccatccatccataataaaacatggaaagtactta 152
DB 88 TGTAGTACTCTTCCTCAAAAGCTGCTTTCCATCCATAAATTAATAAATGGAAGTACTTA 29
QY 153 taagcagagcgacatcatcaagtgtaga 180
DB 28 TGAGCAGAGCGACATATCCAAAGTGAGA 1
XX
RESULT 12
AAS15793/C
ID AAS15793 standard; DNA; 1725 BP.
XX
XX AC AAS15793;
XX
XX 16-JAN-2002 (first entry)
XX
XX Human DNA for Six-Transmembrane Protein of Prostate 1, STMP1.
XX
XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
XX benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
```

```
KW liver cancer; lung cancer; cytostatic; ds.
XX
XX OS Homo sapiens.
XX
XX FH Location/Qualifiers
FT misc_feature 89
FT /*tag= a
FT /*note= "Represents 338 nucleotides of intron 1"
FT misc_feature 162
FT /*tag= b
FT /*note= "Represents 12713 nucleotides of intron 2"
FT CDS 200...1702
FT /*tag= c
FT /*product= "STMP1"
FT misc_feature 697
FT /*tag= d
FT /*note= "Represents 1396 nucleotides of intron 3"
FT misc_feature 1225
FT /*tag= e
FT /*note= "Represents 2372 nucleotides of intron 4"
FT misc_feature 1410
FT /*tag= f
FT /*note= "Represents 2299 nucleotides of intron 5"
XX
XX WO200172962-A2.
XX
XX 04-OCT-2001.
XX
XX 23-MAR-2001; 2001WO-US09410.
XX
XX 24-MAR-2000; 2000US-191929P.
XX (SAAT/) SAATCIOGLU F.
XX Saaticloglu F;
XX WPI; 2001-662926/76.
XX P-PSDB; AAU10187.
XX
XX New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids
PT
XX Example 3; Fig 4B; 114pp; English.
XX
XX The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence encodes a prostate specific protein, Six-Transmembrane
CC Protein of Prostate 1, STMP1.
XX
XX Sequence 1725 BP; 476 A; 340 C; 387 G; 517 T; 5 other;
```

```
Query Match 55.8%; Score 148; DB 22; Length 1725;
Best Local Similarity 100.0%; Pred. No. 8.8e-65;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 33 ccgaattacaattgagggcaaaacaaagagcaagaacaaagtgtgtgtataaaatc 92
DB 1562 CCAGAATTACAAATTGAGGCGCAAAACAAAGAGCAAGAACAAGTTTGGTGTGTATAAAATC 1503
```

CC predicted transmembrane domains, but is encoded by a distinct gene,
 CC localised to chromosome 7q21. STRAP-1 is the prototype member of the
 CC STRAP family of proteins (AAV58194-y58197) which exhibit a high degree
 CC of structural conservation, but which show no significant structural
 CC homology to known human proteins. STRAP-1 is characterised by six
 CC transmembrane domains and intracellular N- and C-termini, suggesting
 CC that it folds in a "serpentine" manner into three extracellular and two
 CC intracellular loops. STRAP-2 exhibits a markedly different mRNA and
 CC protein expression profile relative to STRAP-1, suggesting that these
 CC two STRAP family members are differentially regulated. STRAP-2
 CC expression appears to be very prostate specific, as significant mRNA
 CC expression is not detected in a variety of normal tissues. STRAP-2
 CC expression is downregulated in some prostate cancers, whereas STRAP-1
 CC expression remains at a high level. In non-prostate cancers, STRAP-2
 CC expression is generally absent. The function of the STRAP proteins is not
 CC known. They may be ion channels (from the presence of six transmembrane
 CC domains, a feature which is shared by certain ion channels) or
 CC gap-junction proteins (from immunohistochemical staining). STRAP-1 and
 CC STRAP-2 are cell-surface tumour antigens. Immunisation with a STRAP
 CC protein induces cellular and humoral immune responses against
 CC STRAP-expressing cells. STRAP proteins may be used to identify
 CC specific-binding agents, to produce anticancer vaccines and to generate
 CC specific antibodies. The antibodies may be used for detection, prognosis,
 CC and monitoring of cancers (or susceptibility to cancer), as therapeutic
 CC inhibitors or to target therapeutic agents to their site of action. STRAP
 CC nucleic acids may be used for recombinant protein production, as
 CC cells for screening inhibitors of STRAP expression and for therapeutic
 CC modulation/inhibition of STRAP expression. Since high levels of STRAP
 CC proteins are exposed on the cell surface, they are easily targeted by
 CC systemically administered agents, and because they are expressed mainly
 CC on prostatic epithelial cells, agents targeted to them should have
 CC minimal side effects on other tissues.

XX Sequence 519 BP; 137 A; 105 C; 102 G; 175 T; 0 other;

Query Match 58.1%; Score 154; DB 21; Length 519;
 Best Local Similarity 100.0%; Pred. No. 8.4e-68;
 Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 112 agctcgtttccatcataaaataaaacatggaagtagcttatgagcagagcagacatatcc 171
 Db 519 AGCTCGTGTTCCTCCATAAAFTAAACATGGAAAGTACTTATGACGAGCGCATATCC 460
 QY 172 aagtgtagactgaataaactgaattctctccagtttaaaagcattgctcactgaaggat 231
 Db 459 AAGTGTAGACTGAATAAACTGAATTCTCTCCAGTTTAAAGCATTGCTCACTGAAGGGAT 400
 QY 232 agaagtgtactgccaggaggaggaagtaagccaagg 265
 Db 399 AGAAGTGACTGCCAGGAGGAGGAAGTAAGCCAAGG 366

RESULT 10
 AAS15799/c
 ID AAS15799 standard; DNA; 148 BP.

XX AAS15799;

XX 16-JAN-2002 (first entry)

XX Human Six-Transmembrane Protein of Prostate 1, STMP1, partial exon 6.

XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
 KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
 KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
 KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
 KW liver cancer; lung cancer; cytostatic; ds; exon 6.

OS Homo sapiens.

XX WO200172962-A2.

XX 04-OCT-2001.
 PD 23-MAR-2001; 2001WO-US09410.
 PF 24-MAR-2000; 2000US-191929P.
 XX (SAAT/) SAATCIOGLU F.
 PA Saatioglu F;
 PI WPI; 2001-662926/76.
 XX New polynucleotide for the diagnosis, prevention and treatment for
 XX prostate and testis disorders, particularly prostate cancer, comprises
 XX prostate-specific or testis-specific nucleic acids
 XX Claim 6; Fig 4C; 114pp; English.

XX The invention relates to substantially pure prostate-specific or
 CC testis-specific polypeptides and the nucleic acids encoding them.
 CC Also included are vectors and host cells expressing the proteins, a
 CC transgenic animal expressing the protein, antibodies against the
 CC proteins, probes for detecting the nucleic acids, antisense molecules
 CC for the nucleic acids and methods of isolating modulators of the
 CC proteins. Compounds that modulate the prostate specific or testis
 CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,
 CC cryptorchidism, undescended, retractile, ascending or vanished
 CC testis. Other proliferative disorders for which the modulators may be
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The
 CC present sequence represents partial exon 6 sequence of a prostate
 CC specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.

XX Sequence 148 BP; 36 A; 29 C; 26 G; 57 T; 0 other;

Query Match 55.8%; Score 148; DB 22; Length 148;
 Best Local Similarity 100.0%; Pred. No. 8.9e-65;
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 ccagaataacaattgaggcgcaaacagagcaagacaagaatttgggtgtataaaatc 92

Db 148 CCAGATTACAAATTGAGGGCAAAACAGACGACAGCAACAAAGTTTGGTGTATAAAATC 89

QY 93 tgtagtactcttcctcaaaagctggtttccatccataaaattaaaacatggaagtactta 152

Db 88 TGTAGTACTCTTCTCCAAAGCTCGTTTCCATCCATAAATTAACATGGAAGTACTCTTA 29

QY 153 tgagcagagcgacatatccaagttaga 180

Db 28 TGACGAGCGGACATATCCAAAGTGTAGA 1

RESULT 11
 AAS15808/c
 ID AAS15808 standard; DNA; 148 BP.

XX AAS15808;

XX 16-JAN-2002 (first entry)

XX Human Six-Transmembrane Protein of Prostate 1, STMP1, ORF2 exon 6.

XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
 KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
 KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
 KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
 KW liver cancer; lung cancer; cytostatic; ds; exon 6; ORF2.

XX Homo sapiens.

QY 89 aatctgtagtactctcctcaaaagctcgtttccatccataataaaacatggaagta 148
 |||||
 Db 1406 AATCTGTAGTACTCTTCCCTCAAAAGCTCGTTCCATCCATAAATAAATGGAAGTA 1347
 |||||
 QY 149 cttatgagcagcgacatccaaagttagactg 183
 |||||
 Db 1346 CTTATGAGCAGCGACATATCCAAAGTGTAGACTG 1312
 |||||

RESULT 8
 AAS64300/c
 ID AAS64300 standard; cDNA; 3900 BP.
 XX
 AC AAS64300;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #104.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG00113.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 104; 103pp; English.
 CC
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 3900 BP; 1161 A; 798 C; 892 G; 1049 T; 0 other;

Query Match 58.5%; Score 155; DB 23; Length 3900;

Best Local Similarity 100.0%; Pred. No. 2.6e-68;
 Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 agatccagaattacaaattgagggcaaaacaaagagcaagaacaaagtgtgtgtataa 88
 |||||
 Db 1466 AGATCCAGAATTACAATTGAGGGCAAAACAGAGCAAGCAAAAGTTTGGTGTGTATAA 1407
 |||||
 QY 89 aatctgtagtactctcctcaaaagctcgtttccatccataataaaacatggaagta 148
 |||||
 Db 1406 AATCTGTAGTACTCTTCCCTCAAAAGCTCGTTCCATCCATAAATAAATGGAAGTA 1347
 |||||
 QY 149 cttatgagcagcgacatccaaagttagactg 183
 |||||
 Db 1346 CTTATGAGCAGCGACATATCCAAAGTGTAGACTG 1312
 |||||

RESULT 9
 AAZ49398/c
 ID AAZ49398 standard; cDNA; 519 BP.
 XX
 AC AAZ49398;
 XX
 DT 14-MAR-2000 (first entry)
 XX
 DE Human STRAP-2 partial cDNA.
 XX
 KW Serpentine transmembrane antigen of the prostate; STRAP-2; prostate;
 KW transmembrane domain; type IIIa membrane protein; expression; cancer;
 KW antigen; immunisation; immune response; cellular; humoral;
 KW anticancer vaccine; antibody; detection; diagnosis;
 KW prognosis; monitoring; susceptibility; therapeutic inhibitor;
 KW drug targeting; recombinant protein; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT 1..519
 FT Location/Qualifiers
 FT /tag- a
 FT /partial
 FT /product= "Human STRAP-2 (serpentine transmembrane
 FT antigen of the prostate)"
 FT /note= "No initiation or termination codons given in
 FT the specification"
 XX
 PN WO9962941-A2.
 XX
 PD 09-DEC-1999.
 XX
 PF 01-JUN-1999; 99WO-US12157.
 XX
 PR 01-JUN-1998; 98US-0087520.
 PR 30-JUN-1998; 98US-0091183.
 XX
 PA (UROC-) UROGENESYS INC.
 PA (AFAR/) AFAR D E.
 PA (HUBE/) HUBERT R S.
 PA (LEON/) LEONG K.
 PA (RAIT/) RAITANO A B.
 PA (SAFF/) SAFFRAN D C.
 XX
 PI Afar DE, Hubert RS, Leong K, Raitano AB, Safran DC;
 XX
 DR WPI; 2000-072832/06.
 DR P-PSDB; AAY58195.
 XX
 PT Novel proteins useful as diagnostic markers and therapeutic targets,
 PT particularly for prostatic cancer -
 XX
 PS Claim 12; Fig 9; 83pp; English.
 CC
 CC This sequence represents a partial cDNA encoding a novel human protein,
 CC STRAP-2 (serpentine transmembrane antigen of the prostate). STRAP-2 is
 CC highly homologous to STRAP-1 (AAY58194), particularly throughout the

KW Human: Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 178..1650

XX FT /*tag= a

XX FT /product= "STMP1"

XX WO200172962-A2.

XX PN

XX PD 04-OCT-2001.

XX PF 23-MAR-2001; 2001WO-US09410.

XX PR 24-MAR-2000; 2000US-191929P.

XX PA (SAAT/) SAATCIOGLU F.

XX PI Saatcioglu F;

XX WPI; 2001-662926/76.

XX DR P-PSDB; AAU10187.

XX PT New polynucleotide for the diagnosis, prevention and treatment for

PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids -
XX Claim 5; Fig 4D; 114pp; English.

XX The invention relates to substantially pure prostate-specific or

XX testis-specific polypeptides and the nucleic acids encoding them.

XX Also included are vectors and host cells expressing the proteins, a

XX transgenic animal expressing the protein, antibodies against the

XX proteins, probes for detecting the nucleic acids, antisense molecules

XX for the nucleic acids and methods of isolating modulators of the

XX proteins. Compounds that modulate the prostate specific or testis

XX specific polypeptide are useful to diagnose, prevent or treat disorders

XX of the testis or prostate particularly prostate cancer, benign

XX prostatic hyperplasia, acute prostatitis, testicular cancer,

XX cryptorchidism, undescended, retractile, ascending or vanished

XX testis. Other proliferative disorders for which the modulators may be

XX used include lymphoma, leukaemia, melanoma, ovarian cancer, breast

XX cancer, pancreatic cancer, liver cancer and lung cancer. The

XX present sequence represents the open reading frame of a prostate

XX specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.

XX SQ Sequence 4329 BP; 1315 A; 817 C; 790 G; 1407 T; 0 other;

Query Match 68.7%; Score 182; DB 22; Length 4329;

Best Local Similarity 99.6%; Pred. No. 6,7e-82;

Matches 232; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 33 ccagaattacattgaggcgcaaaacagagcaagcaagaaagtgtgtgtataaaatc 92

Db 1510 CCAGAAATACAAATTTGAGGGGCAAAACAGAGCAAGCAAGAAAGTTGGTGTATATAAAATC 1451

QY 93 tgaagtactctctcaaaagtctgttccatccataaattaaacatggaagtactta 152

Db 1450 TGTAGTACTCTTCTCAAAAGCTGTTTCCATCCATATAATTAACAGTGAAGTACTTA 1391

QY 153 tgaagcagcgacatatccaagtgtagactgaataaaactgaattctctcagtttaag 212

Db 1390 TGAGCAGAGCGACATATCCAAAGTGTAGACTGAATAAAACTGAATTCCTCCAGTTTAAGG 1331

QY 213 catgtctactgaaggatagaagtactgccagaggaggaagtaagcccaagg 265

Db 1330 CATTGCTCACTGAAGGGATAGAAGTGACTGCCAGGAGGGAAGTAAGCCAAG 1278

RESULT 7

AAS76493/C

ID AAS76493 standard; cDNA; 2192 BP.

XX .

AC AAS76493;

XX AC

DT 13-FEB-2002 (first entry)

XX DT

DE DNA encoding novel human diagnostic protein #12297.

XX DE

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX KW

OS Homo sapiens.

XX OS

PN WO200175067-A2.

XX PN

PD 11-OCT-2001.

XX PD

PF 30-MAR-2001; 2001WO-US08631.

XX PF

XX 31-MAR-2000; 2000US-0540217.

XX XX

PR 23-AUG-2000; 2000US-0649167.

XX PR

XX (HYSE-) HYSEQ INC.

XX PA

PI Dmanac RT, Liu C, Tang YT;

XX PI

XX WPI; 2001-639362/73.

XX XX

DR P-PSDB; ABG12306.

XX DR

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

XX Claim 1; SEQ ID No 12297; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AAS64197-AAS94564 represent novel human

CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2192 BP; 559 A; 507 C; 551 G; 575 T; 0 other;

QY 29 agatccagaattacaattgaggcgcaaaacagagcaagcaagtttggtggtataa 88

Db 1466 AGATCCAGAAATTAACAATTTGAGGGGCAAAACAGAGCAAGCAAGTTTGGTGGTATAA 1407

proteins, expressed in human cancers, useful for detecting and treating cancer -
 Claim 4; Fig 9A-9D; 187pp; English.
 The present sequence is human six transmembrane epithelial antigen of the prostate (STEAP)-2 clone GPD3 cDNA. STEAP is a member of cell surface serpentine transmembrane antigens. STEAP-2 gene is located on chromosome 7q21 and is used in gene therapy. Inhibiting the development or progression of a cancer (eg. prostate, colon, bladder, lung, ovarian and pancreatic) expressing STEAP or inhibiting growth or killing cells expressing STEAP in a patient, comprises administering a vaccine composition to the patient. Treating a patient with a cancer that expresses STEAP, or inhibiting growth or killing cells expressing STEAP, comprises administering to the patient a vector encoding single chain monoclonal antibody that comprises the variable domains of the heavy and light chains of the monoclonal antibody that specifically binds to STEAP, such that the vector delivers the single chain monoclonal antibody coding sequence to the cancer cells and the encoded single chain monoclonal antibody is expressed intracellularly.
 Note: The present sequence is also shown in sequence listing of the specification, but it lacks nucleotides at its 5' end.
 Sequence 2453 BP; 648 A; 537 C; 520 G; 748 T; 0 other;
 Query Match 92.1%; Score 244; DB 22; Length 2453;
 Best Local Similarity 100.0%; Pred. No. 4.1e-113;
 Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0
 QY 22 ctgcaaaagatccagataattcaattgaggcgcaaaagagcaagaacaaagtgttggtg 81
 Db 1698 CTCGAAAAGATCCAGAATTACAATTGAGGCGCAAAACAAGACGAACAAGTTGTGG 1639
 QY 82 tgtataaaatctgtactcttctcctcaaaagctogttccatccataataaacaatg 141
 Db 1638 TGTATAAAATCTGTACTCTCTCTCAAAAGCTCGTTTCCATCCATAATTAACAATG 1579
 QY 142 gaaagtacttatgagcagcgacatatccaaagtgtagactgaataaaaactgaattctct 201
 Db 1578 GAAAGTACTTTATCAGCAGAGCGACATATCCAAGTGTAGACTGAATAAACTGAATTCCT 1519
 QY 202 ccagtttaaaagcattctcactgaaggatagaaagtactgccaggagggaagcaagcc 261
 Db 1518 CCAAGTTTAAAGCATTCCTCACTGAAGGGGATAGAGTGACTGCCAGGAGGGAAGTAAGCC 1459
 QY 262 aagg 265
 Db 1458 AAGG 1455
 RESULT 5
 AAS15802/c
 ID AAS15802 standard; cDNA; 1680 BP.
 AC AAS15802;
 XX
 DT 16-JAN-2002 (first entry)
 XX
 DE Human cDNA encoding Six-Transmembrane Protein of Prostate 1, STMP1.
 KW Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer; benign prostatic hyperplasia; acute prostatitis; testicular cancer; cryptorchidism; testicular disorder; proliferative disorder; lymphoma; leukemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer; liver cancer; lung cancer; cytostatic; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 178..1650
 FT /*tag= a
 FT /*product= "STMP1"

RESULT 3
 AAS15810/C
 ID AAS15810 standard; cDNA; 2238 BP.
 XX
 AC AAS15810;
 XX
 DT 16-JAN-2002 (first entry)
 XX
 DE Human ORF2 of Six-Transmembrane Protein of Prostate 1, STMP1.
 XX
 KW Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
 KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
 KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
 KW leukemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
 KW liver cancer; lung cancer; cytostatic; ss; ORF2.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 188..1552
 FT /*tag= a
 FT /product= "STMP1, ORF2"
 XX
 PN WO200172962-A2.
 XX
 PD 04-OCT-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09410.
 XX
 PR 24-MAR-2000; 2000US-191929P.
 XX
 PA (SAAT/) SAATCIOGLU F.
 XX
 PI Saatioglu F;
 XX
 DR WPI; 2001-662926/76.
 DR P-PSDB; AAU10188.
 XX
 PT New polynucleotide for the diagnosis, prevention and treatment for
 PT prostate and testis disorders, particularly prostate cancer, comprises
 PT prostate-specific or testis-specific nucleic acids -
 XX
 PS Claim 5; Fig 4G; 114pp; English.
 XX
 CC The invention relates to substantially pure prostate-specific or
 CC testis-specific polypeptides and the nucleic acids encoding them.
 CC Also included are vectors and host cells expressing the proteins, a
 CC transgenic animal expressing the protein, antibodies against the
 CC proteins, probes for detecting the nucleic acids, antisense molecules
 CC for the nucleic acids and methods of isolating modulators of the
 CC proteins. Compounds that modulate the prostate specific or testis
 CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,
 CC cryptorchidism, undescended, retractile, ascending or vanished
 CC testis. Other proliferative disorders for which the modulators may be
 CC used include lymphoma, leukemia, melanoma, ovarian cancer, breast
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The
 CC present sequence represents the second open reading frame of a prostate
 CC specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.
 XX
 SQ Sequence 2238 BP; 607 A; 457 C; 453 G; 721 T; 0 other;
 XX
 Query Match 92.1%; Score 244; DB 22; Length 2238;
 Best Local Similarity 100.0%; Pred. No. 4.1e-113;
 Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 22 ctgcacaaagatccgaattcaattgaggcgcaaacagagcgacgaacaaagtgtgtg 81
 Db 1531 CTGCAAAAGATCCAGATTCAATTGAGGGCAAAACAGACGACAAAGATTGTGG 1472

QY 82 tgtataaatctgtagtactcttctcaaaagctcggttccatccataataaaacatg 141
 Db 1471 TGTATAAAATCTGTAGTACTCTTCTCTCAAAAGCTCGTTCATCCATATAAAACATG 1412
 QY 142 gaaagtacttatgagcagagcgacatatccaaagtgtagactgaataaaactgaattctct 201
 Db 1411 GAAAGTACTTATGAGCAGAGCGACATATCCAAGTGTAGACTGAATAAAACTGAATTCCT 1352
 QY 202 ccagtttaagcattgtcactgaaggatagaaagtgcactccaggagggaagtaagcc 261
 Db 1351 CCAGTTTAAAGCATGTGCTCACTGAAGGATAGAAAGTACTCCAGGAGGAAAGTAAGCC 1292
 QY 262 aagg 265
 Db 1291 AAGG 1288
 RESULT 4
 AAD07072/C
 ID AAD07072 standard; cDNA; 2453 BP.
 XX
 AC AAD07072;
 XX
 DT 06-AUG-2001 (first entry)
 XX
 DE Human six transmembrane epithelial antigen of prostate-2 clone GTD3 CDNA.
 XX
 KW Human; cytostatic; antiproliferative; vaccine; gene therapy;
 KW six transmembrane epithelial antigen of the prostate-2; STEAP-2;
 KW chromosome 7q21; cancer; prostate; colon; bladder; lung; ovarian;
 KW pancreatic; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT RBS 352..358
 FT /*tag= a
 FT /note= "kozak region"
 FT CDS 355..1719
 FT /*tag= b
 FT /product= "Human six transmembrane epithelial antigen
 FT of the prostate (STEAP)-2"
 FT 709..2073
 FT /*tag= c
 FT /product= "Human six transmembrane epithelial antigen
 FT of the prostate (STEAP)-2, alternative version"
 FT /note= "CDS does not include start and stop codon"
 FT /transl_except= (pos:1714..1722, aa:Asp-Ala)
 FT /transl_except= (pos:1834..1842, aa:Arg-Ser)
 FT /transl_except= (pos:1957..1965, aa:Glu-Gly)
 FT /transl_except= (pos:2050..2058, aa:Thr-Ser)
 FT /transl_except= (pos:2062..2070, aa:Asn-Phe)
 FT /note= "Inframe stop codon alters the reading frame"
 FT /partial
 XX
 PN WO200140276-A2.
 XX
 PD 07-JUN-2001.
 XX
 PF 06-DEC-2000; 2000WO-US33040.
 XX
 PR 06-DEC-1999; 99US-0455486.
 XX
 PA (UROG-) UROGENESYS INC.
 XX
 PI Afar DEH, Hubert RS, Raitano AB, Safran DC, Mitchell SC, Faris M;
 PI Jakobovits A;
 XX
 DR WPI; 2001-367804/38.
 DR P-PSDB; AAE02781, AAE02841.
 XX
 PT New STEAP (six transmembrane epithelial antigen of the prostate)

XX (INCY-) INCYTE PHARM INC.
XX Lal P, Guegler KJ, Corley NC;
XX WPI: 2000-062671/05.
XX P-PSDB; AAY52589.
XX New human prostate growth-associated membrane proteins, for treating or
XX preventing cancer and reproductive disorders
XX Claim 7; Page 67; 72pp; English.
XX This sequence represents cDNA encoding human prostate growth-associated
XX protein PGAMP-1. Nucleotides encoding PGAMP-1 were initially identified
XX in a prostate cDNA library, this sequence representing a consensus.
XX Human prostate growth associated membrane proteins PGAMP-1 and PGAMP-2
XX (AAY52590) may be used to raise specific antibodies and to screen for
XX specific modulators (agonists, antagonists or other potential
XX therapeutic agents). Antagonists of PGAMP are used to treat or prevent a
XX wide range of cancers (solid tumours, leukaemia, lymphoma etc.) and
XX reproductive disorders (such as infertility, endometriosis, polycystic
XX ovarian syndrome, prostatitis). PGAMP-encoding nucleic acids, its
XX fragments and complements, may be used for recombinant production of
XX PGAMP proteins, in gene therapy (e.g., as antisense molecules, triplex-
XX forming molecules and ribozymes), and as diagnostic probes and primers.
XX Anti-PGAMP antibodies may be used for diagnosis and monitoring of
XX PGAMP-related diseases by standard immunoassays, as therapeutic
XX antagonists (including targeted delivery of other drugs), and in
XX competitive drug screens.
XX Sequence 1213 BP; 335 A; 239 C; 215 G; 424 T; 0 other;

Query Match 92.1%; Score 244; DB 21; Length 1213;
Best Local Similarity 100.0%; Pred. No. 4.le-113;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 22 ctgcaaaagatccagaattacaattgaggcgcaaaacagagcaagaacaaagtgttggtg 81
Db 487 CTGCAAAAGATCCAGAAATACAAATTTGAGGGCAAAACAGAGCAAGCAAAAGTTGGTGG 428
Qy 82 tqtataaaatctgtactctctcaaaagctgttccatccataaaataaaacatg 141
Db 427 TGTATAAAATCTGTAGTACTCTTCTCAAAAGCTGTTCCATCAATTAATAAACATG 368
Qy 142 gaaagtacttatgacagagcacatatccaaagttagactgaataaaactgaattctct 201
Db 367 GAAAGTACTTATGACGAGCGACATATCCAAAGTGTAGACTGAATAAAACTCAATTCTCT 308
Qy 202 ccagtttaaaagcattgctcactgaaggagatagaagtactgacagagggaagtaagcc 261
Db 307 CCAGTTTAAAGCATTGCTCACTGAAGGGATAGAGTACTGCCAGGAGGAAAGTAAGCC 248
Qy 262 aagg 265
Db 247 AAGG 244

RESULT 2
AAS15811/C
ID AAS15811 standard; cDNA; 1561 BP.
XX AAS15811;
XX
XX 16-JAN-2002 (first entry)
XX Human cDNA encoding ORF2 of Six-Transmembrane Protein of Prostate 1.
XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
XX benign prostatic hyperplasia; acute prostatitis; testicular cancer;
XX cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
XX leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;

KW liver cancer; lung cancer; cytostatic; ss; ORF2.
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 188..1552
XX FT /*tag= a
XX FT /product= "STMP1, ORF2"
XX WO200172962-A2.
XX PD 04-OCT-2001.
XX PF 23-MAR-2001; 2001WO-US09410.
XX PR 24-MAR-2000; 2000US-191929P.
XX PA (SAAT/) SAATCIOGLU F.
XX PI Saatioglu F;
XX WPI: 2001-662926/76.
XX P-PSDB; AAU10188.
XX New polynucleotide for the diagnosis, prevention and treatment for
XX prostate and testis disorders, particularly prostate cancer, comprises
XX prostate-specific or testis-specific nucleic acids
XX Claim 4; Fig 4H; 114pp; English.
XX The invention relates to substantially pure prostate-specific or
XX testis-specific polypeptides and the nucleic acids encoding them.
XX Also included are vectors and host cells expressing the proteins, a
XX transgenic animal expressing the protein, antibodies against the
XX proteins, probes for detecting the nucleic acids, antisense molecules
XX for the nucleic acids and methods of isolating modulators of the
XX proteins. Compounds that modulate the prostate specific or testis
XX specific polypeptide are useful to diagnose, prevent or treat disorders
XX of the testis or prostate particularly prostate cancer, benign
XX prostatic hyperplasia, acute prostatitis, testicular cancer,
XX cryptorchidism, undescended, retractile, ascending or vanished
XX testis. Other proliferative disorders for which the modulators may be
XX used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
XX cancer, pancreatic cancer, liver cancer and lung cancer. The
XX present sequence encodes prostate specific protein, Six-Transmembrane
XX Protein of Prostate 1, STMP1, ORF2.
XX Sequence 1561 BP; 420 A; 319 C; 346 G; 476 T; 0 other;

Query Match 92.1%; Score 244; DB 22; Length 1561;
Best Local Similarity 100.0%; Pred. No. 4.le-113;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 22 ctgcaaaagatccagaattacaattgaggcgcaaaacagagcaagaacaaagtgttggtg 81
Db 1531 CTGCAAAAGATCCAGAAATACAAATTTGAGGGCAAAACAGAGCAAGCAAAAGTTGGTGG 1472
Qy 82 tqtataaaatctgtactctctcaaaagctgttccatccataaaataaaacatg 141
Db 1471 TGTATAAAATCTGTAGTACTCTTCTCAAAAGCTGTTCCATCAATTAATAAACATG 1412
Qy 142 gaaagtacttatgacagagcacatatccaaagttagactgaataaaactgaattctct 201
Db 1411 GAAAGTACTTATGACGAGCGACATATCCAAAGTGTAGACTGAATAAAACTCAATTCTCT 1352
Qy 202 ccagtttaaaagcattgctcactgaaggagatagaagtactgacagagggaagtaagcc 261
Db 1351 CCAGTTTAAAGCATTGCTCACTGAAGGGATAGAGTACTGCCAGGAGGAAAGTAAGCC 1292
Qy 262 aagg 265
Db 1291 AAGG 1288

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:59:54 ; Search time 600.75 seconds
(without alignments)
757.357 Million cell updates/sec

Title: US-09-802-520-7
Perfect score: 265
Sequence: 1 ctgactctgtgtatctgcaaa.....ggagggaagtaagccaagg 265

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	244	92.1	1213	21	AAZ46296
C 2	244	92.1	1561	22	AAS15811
C 3	244	92.1	2238	22	AAS15810
C 4	244	92.1	2453	22	AAD07072
C 5	182	68.7	1680	22	AAS15802
C 6	182	68.7	4329	22	AAS15801
C 7	155	58.5	2192	23	AAS76493
C 8	155	58.5	3900	23	AAS64300
C 9	154	58.1	519	21	AAZ49398

C 10	148	55.8	148	22	AAS15799	Human Six-Transmem
C 11	148	55.8	148	22	AAS15808	Human Six-Transmem
C 12	148	55.8	1725	22	AAS15793	Human DNA for Six-
C 13	93	35.1	250	19	AAV51347	Reverse DNA sequen
C 14	93	35.1	250	19	AAV58683	Prostate tumour sp
C 15	93	35.1	250	21	AAO06446	Human immunogenic
C 16	93	35.1	250	22	AAS63654	Human prostate cDN
C 17	93	35.1	250	22	AAS10205	Human prostate tum
C 18	93	35.1	250	22	AAS10205	Human prostate-spe
C 19	93	35.1	250	22	AAS10205	Human prostate-spe
C 20	93	35.1	250	22	AAS10205	Human prostate-spe
C 21	86	32.5	366	19	AAV51349	Reverse DNA sequen
C 22	86	32.5	366	19	AAV58685	Prostate tumour sp
C 23	86	32.5	366	21	AAO06448	Human immunogenic
C 24	86	32.5	366	22	AAS63656	Human prostate cDN
C 25	86	32.5	366	22	AAS10207	Human prostate tum
C 26	86	32.5	366	22	AAS93564	Human prostate-spe
C 27	86	32.5	366	22	AAH84878	Human prostate-spe
C 28	86	32.5	366	22	AAH02629	Prostate tumour an
C 29	86	32.5	366	22	AAS15798	Human Six-Transmem
C 30	85	32.1	165	22	AAS15807	Human Six-Transmem
C 31	42	15.8	2102	22	AAS15813	Human cDNA encodin
C 32	21	7.9	183	21	AAZ49400	Human STRAP-2 gene
C 33	21	7.9	183	22	AAD07075	SSH fragment cDNA
C 34	21	7.9	301	21	AAO06521	Human immunogenic
C 35	21	7.9	301	22	AAS63729	Human prostate cDN
C 36	21	7.9	301	22	AAS93637	Human prostate-spe
C 37	21	7.9	301	22	AAH84951	Human prostate-spe
C 38	21	7.9	301	22	AAH02702	Prostate tumour an
C 39	21	7.9	718	22	AAS15809	Human Six-Transmem
C 40	21	7.9	1255	22	AAE72762	Human prostate can
C 41	21	7.9	1325	21	AAF15730	Human prostate can
C 42	21	7.9	1325	21	AAE72787	Human prostate can
C 43	18	6.8	415	22	AAZ23820	Human breast cance
C 44	18	6.8	568	22	AAH14967	Human breast cance
C 45	18	6.8	761	22	AAI96249	Human neuroblastom

ALIGNMENTS

RESULT 1

AAZ46296/c
ID AAZ46296 standard; cDNA; 1213 BP.

XX AC AAZ46296;

XX DT 07-MAR-2000 (first entry)

XX DE Human prostate growth-associated membrane protein PGAMP-1 cDNA.

XX DE Prostate growth-associated membrane protein; PGAMP-1; prostate;

XX DE consensus; antibody; screening; modulator; agonist; antagonist;

XX DE therapeutic agent; cancer; solid tumour; leukaemia; lymphoma;

XX DE reproductive disorder; infertility; endometriosis;

XX DE polycystic ovarian syndrome; prostatitis; recombinant expression;

XX DE gene therapy; antisense therapy; ribozyme; diagnosis; diagnosis;

XX DE monitoring; immunoassay; targeting; drug delivery; drug screening; ds.

XX OS Homo sapiens.

XX FH Key

XX FT CDS

XX FT Location/Qualifiers

XX FT 83..508

XX FT /**tag= a

XX FT /product= "Human PGAMP-1"

XX FT WO961469-A2.

XX FT 02-DEC-1999.

XX FT 17-MAY-1999;

XX FT 99WO-US10888.

XX FT 22-MAY-1998;

XX FT 98US-0083521.


```

----- Project Information -----
Center project name: ccy
Center clone name: 167P22

----- Summary Statistics -----
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171178 bases at least Q40
Consensus quality: 171910 bases at least Q30
Consensus quality: 172236 bases at least Q20
Insert size: 130000; agarose-fp
Insert size: 172615; sum-of-ctligs
Quality coverage: 12.58x in Q20 bases; sum-of-ctligs
Quality coverage: 9.47x in Q20 bases; sum-of-ctligs

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2438: contig of 2438 bp in length
* 2439 2538: gap of unknown length
* 2539 8133: contig of 5595 bp in length
* 8134 8233: gap of unknown length
* 8234 40378: contig of 32145 bp in length
* 40379 40478: gap of unknown length
* 40479 172915: contig of 132437 bp in length.
*
* Location/Qualifiers
* 1. 172915
* /organism="Papio cynocephalus anubis"
* /db_xref="taxon:9555"
* /clone="RP41-167P22"
* /clone_1fb="RP41"
* 1. 2438
* /note="assembly_fragment"
* clone_end:SP6
* vector_side:right"
* 2539. 8133
* /note="assembly_fragment"
* 8234. 40378
* /note="assembly_fragment"
* clone_end:T7
* vector_side:right"
* 40479. 172915
* /note="assembly_fragment"
* 52509 a 31733 c 32277 g 56096 t 300 others
BASE COUNT
ORIGIN

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Afar,D.E., Hubert,R.S., Raitano,A.B., Saffran,D.C., Mitchell,S.C.,
TITLE Serpentine transmembrane antigens expressed in human prostate
JOURNAL Patent: WO 0140276-A 12 07-JUN-2001;
Urogenesys, Inc. (US)
FEATURES
source
1. 183
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 40 a 41 c 43 g 59 t
ORIGIN

Query Match 7.9%; Score 21; DB 6; Length 183;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcaagctgggtatctgcaaa 21
|||||
Db 29 CTCAGCTCGGTATCTGCCAA 9

RESULT 14
AX106507 301 bp DNA linear PAT 30-APR-2001
LOCUS
DEFINITION Sequence 288 from Patent WO0125272.
ACCESSION AX106507
VERSION AX106507.1 GI:13922186
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Xu,J., Skeiky,Y.A., Reed S.G. and Cheever M.A.
TITLE Compositions and methods for therapy and diagnosis of prostate
JOURNAL Patent: WO 0125272-A 288 12-APR-2001;
CORIXA CORPORATION (US)
FEATURES
source
1. 301
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 108 a 49 c 61 g 83 t
ORIGIN

Query Match 7.9%; Score 21; DB 6; Length 301;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctcaagctgggtatctgcaaa 21
|||||
Db 91 CTCAGCTCGGTATCTGCCAA 111

RESULT 15
AX140798 301 bp DNA linear PAT 31-MAY-2001
LOCUS
DEFINITION Sequence 288 from Patent WO0134802.
ACCESSION AX140798
VERSION AX140798.1 GI:14280909
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 301)

```

AUTHORS

Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,
Skelky,Y.A. and Wang,A.

TITLE

Compositions and methods for the therapy and diagnosis of prostate

JOURNAL

Cancer
Patent: WO 0134802-A 288 17-MAY-2001;

FEATURES

CORIXA CORPORATION (US)
Location/Qualifiers
1. 301
/organism="Homo sapiens"
/db_xref="taxon:9606"

source

1. 301
/db_xref="taxon:9606"

BASE COUNT

108 a 49 c 61 g 83 t

ORIGIN

Query Match

7.9% Score 21; DB 6; Length 301;

Best Local Similarity 100.0%; Pred.No. 1;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctcagctctgtatctgcaaa 21
|||||

Db 91 CTCAGCTCTGTATCTGC AAA 111

Search completed: September 20, 2002, 09:55:54
Job time: 13212 sec

REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
BASE COUNT
ORIGIN

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 366)
Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
Reed, S.G., Kalos, M.D., Retter, M.W., Stolk, J.A., Day, C.H.,
Skeiky, Y.A. and Wang, A.
Compositions and methods for the therapy and diagnosis of prostate
cancer.
Patent: WO 0134802-A 215 17-MAY-2001;
CORIXA CORPORATION (US)
Location/Qualifiers
1. .366
/organism="Homo sapiens"
/db_xref="taxon:9606"
119 a 79 c 69 g 96 t 3 others

Query Match 32.5%; Score 86; DB 6; Length 366;
Best Local Similarity 100.0%; Pred. No. 8.3e-38;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 actgaataaaactgaattctctcagtttaaacgattgctcactgaaggatagaagtga 239
|||||
Db 33 ACTGAATAAACTGAATCTCTCCAGTTTAAAGCATTCCTCACTGAAGGATAGAAGTGA 92
|||||

QY 240 ctgccaggagggaagtaagccaagg 265
|||||
Db 93 CTGCCAGGAGGGAAGTAAGCCAAGG 118
|||||

RESULT 10
LOCUS AX200585
DEFINITION Sequence 215 from Patent WO0151633.
ACCESSION AX200585
VERSION AX200585.1 GI:15390405
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 366)
Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
Reed, S.G., Kalos, M.D., Fanger, G.R., Day, C.H., Retter, M.W.,
Stolk, J.A., Skeiky, Y.A., Wang, A. and Meagher, M.J.
Compositions and methods for the therapy and diagnosis of prostate
cancer.
Patent: WO 0151633-A 215 19-JUL-2001;
CORIXA CORPORATION (US)
Location/Qualifiers
1. .366
/organism="Homo sapiens"
/db_xref="taxon:9606"
119 a 79 c 69 g 96 t 3 others

Query Match 32.5%; Score 86; DB 6; Length 366;
Best Local Similarity 100.0%; Pred. No. 8.3e-38;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 actgaataaaactgaattctctcagtttaaacgattgctcactgaaggatagaagtga 239
|||||
Db 33 ACTGAATAAACTGAATCTCTCCAGTTTAAAGCATTCCTCACTGAAGGATAGAAGTGA 92
|||||

QY 240 ctgccaggagggaagtaagccaagg 265
|||||
Db 93 CTGCCAGGAGGGAAGTAAGCCAAGG 118
|||||

RESULT 11
LOCUS AX267241
DEFINITION Sequence 215 from Patent WO0173032.
ACCESSION AX267241
VERSION AX267241.1 GI:16516019
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
Kalos, M.D., Fanger, G.R., Retter, M.W., Stolk, J.A., Day, C.H.,
Vedvick, T.S., Carter, D., Li, S.X., Wang, A., Skeiky, Y.A., Hepler, W.T.
and Henderson, R.A.
Compositions and methods for the therapy and diagnosis of prostate
cancer.
Patent: WO 0173032-A 215 04-OCT-2001;
CORIXA CORPORATION (US)
Location/Qualifiers
1. .366
/organism="Homo sapiens"
/db_xref="taxon:9606"
119 a 79 c 69 g 96 t 3 others

Query Match 32.5%; Score 86; DB 6; Length 366;
Best Local Similarity 100.0%; Pred. No. 8.3e-38;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 actgaataaaactgaattctctcagtttaaacgattgctcactgaaggatagaagtga 239
|||||
Db 33 ACTGAATAAACTGAATCTCTCCAGTTTAAAGCATTCCTCACTGAAGGATAGAAGTGA 92
|||||

QY 240 ctgccaggagggaagtaagccaagg 265
|||||
Db 93 CTGCCAGGAGGGAAGTAAGCCAAGG 118
|||||

RESULT 12
LOCUS AC099742
DEFINITION Papio cynocephalus anubis clone RP41-167P22, WORKING DRAFT
SEQUENCE, 4 unordered pieces.
ACCESSION AC099742
VERSION AC099742.1 GI:17017546
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE olive baboon.
ORGANISM Papio cynocephalus anubis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Cercopithecidae;
Cercopithecinae; Papio.
1 (bases 1 to 172915)
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-O.,
Legaspi, R., Maduro, Q.B., Maduro, V.B., Masiello, C., Mastrian, S.D.,
McGloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y.,
Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
Tsaurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Zhang, L.-H. and Green, E.D.
Unpublished
NISC Comparative Sequencing Initiative
2 (bases 1 to 172915)
Green, E.D.
Direct Submission
Submitted (20-NOV-2001) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact nisc_mouse@nhgri.nih.gov

JOURNAL Patent: WO 0173032-A 213 04-OCT-2001;
CORIXA CORPORATION (US)

FEATURES
source
Location/Qualifiers
1..250
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 81 a 53 c 42 g 69 t 5 others
ORIGIN

Query Match 35.1%; Score 93; DB 6; Length 250;
Best Local Similarity 100.0%; Pred. No. 8.5e-42;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 agttagactgaataaaactgaattctctccagtttaagcattgctcactgaaggata 232
|||||
Db 26 AGTGTAGACTGAATAAACTGAATTCCTCCAGTTTAAAGCATTGCTCACTGAAGGATA 85

QY 233 gaagtactgccaggagggaagtaagccaagg 265
|||||
Db 86 GAAGTACTGCCAGGAGGAAAGTAAGCCAAGG 118

RESULT 8

AX106434
LOCUS AX106434 366 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 215 from Patent WO0125272.
ACCESSION AX106434
VERSION AX106434.1 GI:13922113

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 366)

AUTHORS Xu,J., Skelky,Y.A., Reed,S.G. and Cheever,M.A.

TITLE Compositions and methods for therapy and diagnosis of prostate cancer

JOURNAL Patent: WO 0125272-A 215 12-APR-2001;

CORIXA CORPORATION (US)

FEATURES
source
Location/Qualifiers
1..366
/organism="Homo sapiens"
/db_xref="taxon:9606"

misc_feature 1..366
/note="A, T, C or G"

BASE COUNT 119 a 79 c 69 g 96 t 3 others

ORIGIN

Query Match 32.5%; Score 86; DB 6; Length 366;
Best Local Similarity 100.0%; Pred. No. 8.3e-38;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 actgaataaaactgaattctctccagtttaagcattgctcactgaaggataagtgata 239
|||||
Db 33 ACTGATTAAGTGAATTCCTCCAGTTTAAAGCATTGCTCACTGAGGGATAGAGTGA 92

QY 240 ctgccaggagggaagtaagccaagg 265
|||||
Db 93 CTGCCAGGAGGGAAGTAAGCCAAGG 118

RESULT 9

AX140725
LOCUS AX140725 366 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 215 from Patent WO0134802.
ACCESSION AX140725
VERSION AX140725.1 GI:14280836

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 366)

AUTHORS Xu,J., Skelky,Y.A., Reed,S.G. and Cheever,M.A.

TITLE Compositions and methods for therapy and diagnosis of prostate cancer

JOURNAL Patent: WO 0125272-A 215 12-APR-2001;

CORIXA CORPORATION (US)

FEATURES
source
Location/Qualifiers
1..366
/organism="Homo sapiens"
/db_xref="taxon:9606"

Query Match 35.1%; Score 93; DB 6; Length 250;
Best Local Similarity 100.0%; Pred. No. 8.5e-42;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 agttagactgaataaaactgaattctctccagtttaagcattgctcactgaaggata 232
|||||
Db 26 AGTGTAGACTGAATAAACTGAATTCCTCCAGTTTAAAGCATTGCTCACTGAAGGATA 85

QY 233 gaagtactgccaggagggaagtaagccaagg 265
|||||
Db 86 GAAGTACTGCCAGGAGGAAAGTAAGCCAAGG 118

RESULT 6

AX200583
LOCUS AX200583 250 bp DNA linear PAT 29-AUG-2001
DEFINITION Sequence 213 from Patent WO0151633.
ACCESSION AX200583
VERSION AX200583.1 GI:15390403

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 250)

AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,
Stolk,J.A., Skelky,Y.A., Wang,A. and Meagher,M.J.

TITLE Compositions and methods for the therapy and diagnosis of prostate cancer

JOURNAL Patent: WO 0151633-A 213 19-JUL-2001;

CORIXA CORPORATION (US)

FEATURES
source
Location/Qualifiers
1..250
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 81 a 53 c 42 g 69 t 5 others

ORIGIN

Query Match 35.1%; Score 93; DB 6; Length 250;
Best Local Similarity 100.0%; Pred. No. 8.5e-42;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 agttagactgaataaaactgaattctctccagtttaagcattgctcactgaaggata 232
|||||
Db 26 AGTGTAGACTGAATAAACTGAATTCCTCCAGTTTAAAGCATTGCTCACTGAAGGATA 85

QY 233 gaagtactgccaggagggaagtaagccaagg 265
|||||
Db 86 GAAGTACTGCCAGGAGGGAAGTAAGCCAAGG 118

RESULT 7

AX267239
LOCUS AX267239 250 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 213 from Patent WO0173032.
ACCESSION AX267239
VERSION AX267239.1 GI:16516017

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (sites)

AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Kalos,M.D., Fanger,G.R., Retter,M.W., Stolk,J.A., Day,C.H.,
Vedvick,T.S., Carter,D., Li,S.X., Wang,A., Skelky,Y.A., Hepler,W.T.
and Henderson,R.A.

TITLE Compositions and methods for the therapy and diagnosis of prostate cancer

JOURNAL Patent: WO 0173032-A 213 04-OCT-2001;

CORIXA CORPORATION (US)

FEATURES
source
Location/Qualifiers
1..250
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 81 a 53 c 42 g 69 t 5 others

ORIGIN

Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: http://www.nisc.nih.gov
 Contact: nisc_mouse@hgrl.nih.gov
 ----- Project Information
 Center project name: ces
 Center clone name: 120K11
 ----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 161001 bases at least Q40
 Consensus quality: 161422 bases at least Q30
 Consensus quality: 161562 bases at least Q20
 Insert size: 143000; agarose-fp
 Insert size: 162428; sum-of-contigs
 Quality coverage: 10.76x in Q20 bases; agarose-fp
 Quality coverage: 9.47x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 13922: contig of 13922 bp in length
 13923 14022: gap of unknown length
 14023 36248: contig of 22226 bp in length
 36249 36348: gap of unknown length
 36349 66192: contig of 29844 bp in length
 66193 66292: gap of unknown length
 66293 92168: contig of 25876 bp in length
 92169 92269: gap of unknown length
 92269 122036: contig of 29768 bp in length
 122037 122136: gap of unknown length
 122137 162928: contig of 40792 bp in length.

FEATURES

source
 1. .162928
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="RP43-120K11"
 /clone_lib="RP43"
 misc_feature
 1. .13922
 /note="assembly_fragment"
 clone_end:T7
 vector_side:left
 misc_feature
 14023 - 36248
 /note="assembly_fragment"
 misc_feature
 36349 - 66192
 /note="assembly_fragment"
 misc_feature
 66293 - 92168
 /note="assembly_fragment"
 misc_feature
 92269 - 122036
 /note="assembly_fragment"
 misc_feature
 122137 - 162928
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:right
 BASE COUNT 51528 a 29878 c 30533 g 50485 t 504 others
 ORIGIN

Query Match 37.4%; Score 99; DB 2; Length 162928;
 Best Local Similarity 99.3%; Pred. No. 2.6e-45;
 Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 33 ccagaattacattgaggccaaacagagcaagaattgtgtgtatataaatc 92
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 116287 CCAGAAATTCAGTTGAGGCGCAAAACAGAGCAAGAAAGTTGGTGTGTATAAATC 116346

Qy 93 tgtagtactctcctcaaaagctggtttccatccataataataacatggaagtactta 152

^ ^ ^

||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 116347 TGTAGTACTCTTCTTCAAAGCTGTTCCATCCATAAATTAACATGGAAGTACTTA 116406
 ||||||||||||||||||||||||||||||||||||||||||||||||
 Qy 153 tgcagcagcgacacatcccaagtgtagact 182
 ||||||||||||||||||||||||||||||||||||||||||||||||
 Db 116407 TGACAGAGCGACATATCCAAAGTGTAGACT 116436

RESULT 4
 AX106432
 LOCUS AX106432 250 bp DNA linear PAT 30-APR-2001
 DEFINITION Sequence 213 from Patent WO0125272.
 ACCESSION AX106432
 VERSION AX106432.1 GI:13922111
 KEYWORDS
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 250)

AUTHORS Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.
 TITLE Compositions and Methods for therapy and diagnosis of prostate cancer

JOURNAL Patent: WO 0125272-A 213 12-APR-2001;

FEATURES
 Location/Qualifiers

source
 1. .250
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

misc_feature
 1. .250

/note="n = A,T,C or G"

BASE COUNT 81 a 53 c 42 g 69 t 5 others

ORIGIN

Query Match 35.1%; Score 93; DB 6; Length 250;
 Best Local Similarity 100.0%; Pred. No. 8.5e-42;
 Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 173 agtgtagactgaataaaactgaattctctccagtttaagcattgtctcactgaaggata 232
 ||||||||||||||||||||||||||||||||||||||||||||||||
 Db 26 AGTGAGTACTGAATAAACTGAATCTCTCCAGTTAAAGCATTCCTCACTGAGGGATA 85

Qy 233 gaagtgaactccaggaggagaaagtaagccaagg 265
 ||||||||||||||||||||||||||||||||||||||||||||||||
 Db 86 GAAGTGAAGTCCAGGAGGAAAGTAAGCCAGG 118

RESULT 5

AX140723
 LOCUS AX140723 250 bp DNA linear PAT 31-MAY-2001
 DEFINITION Sequence 213 from Patent WO0134802.
 ACCESSION AX140723
 VERSION AX140723.1 GI:14280834
 KEYWORDS
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 250)

AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
 Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,
 Skeiky,Y.A. and Wang,A.

TITLE Compositions and Methods for the therapy and diagnosis of prostate cancer

JOURNAL Patent: WO 0134802-A 213 17-MAY-2001;

FEATURES
 Location/Qualifiers

source
 1. .250
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 81 a 53 c 42 g 69 t 5 others

ORIGIN

COMMENT

/note="unnamed protein product"

/codon_start=1
/protein_id="CAC42678.1"
/db_xref="GI:14536714"
/translation="MESTSMGSPKSLSETCLPNCINGIKDKARKVTGCVIGSDPFAKS
LTIRLRCGVHVIVGSRNPKFASFEFPHVDVTHEDALTKTNILEVAHREHYTSLW
DRLHLVGLKILIDVSNRRINQINQFESNAYLASFLPDLSLIYKGVNVSAAWALQGPKD
ASRQYICSNIIQARQVIELARQNFIPIDGLSSAREIENPLRFTLWRGPVVV
AIASTAEFLYSFVRDVIHPYARNOQSDFYKPIEIVNKTPIPIVAITLLSLVYLAGLL
AAAYOLGYTYRFPWLEWLCQKQLGLSLFFFAVHVWYSLCLPMRRESRYLFL
NNAYQQVHANIESNWEVEVRIEMYSFGIMSLGLSLAVTSPSVSNALNNREFS
FIQSTGLGVALLISTEHLVLYGWKRAPEEYRYFTTPNFVLALVLPISVILDLQLC
RYPD"

BASE COUNT 648 a 537 c 520 g 748 t

ORIGIN

Query Match 92.1%; Score 244; DB 6; Length 2453;

Best Local Similarity 100.0%; Pred. No. 5.5e-128; Indels 0; Gaps 0;

Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 ctgcaaaagatccagaattacattgaggcgaaacaaagcaagcaaaagtgttggtg 81
|||||
Db 1698 CFGCAAAAGATCCAGAAATACAAATGAGGGCAAAACAGAGCAAGCAAAAGTTGGTGG 1639
|||||
QY 82 tqtataaatctgtgactctctcctcaaaagctgttccatccataataaaacatg 141
|||||
Db 1638 TGTATAAAATCTGTAGTACTCTCTCTCAAAAGCTGTTTCCATCCATAAATTAACAATG 1579
|||||
QY 142 gaaagtactatgacagagcacatatccaagtataactgaataaaactgaattctct 201
|||||
Db 1578 GAAGTACTATTAGCAGAGGCACATATCCAGTGTAGTACTGAATAAACTGAATCTCT 1519
|||||
QY 202 ccagtttaagcattgctcaactgaaggatagaagtgactgctgaggaggaagtaagcc 261
|||||
Db 1518 CCAGTTTAAAGCATTCCTCAGTGAAGGGATAGAGTACTGCCAGAGGAGGAAAGTAAAGCC 1459
|||||
QY 262 aagg 265
|||||
Db 1458 AAGG 1455

RESULT 2
HSAC002064/C 156214 bp DNA linear PRI 09-MAY-1997
LOCUS Human BAC clone RG016J04 from 7q21, complete sequence.

DEFINITION AC002064
ACCESSION AC002064
VERSION AC002064.1 GI:2076723
KEYWORDS HTG.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 156214)
AUTHORS Gattung,S. and Maggi,L.
TITLE The sequence of H. sapiens BAC clone RG016J04
JOURNAL Unpublished (1997)

REFERENCE 2 (bases 1 to 156214)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1997)

Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA
http://genome.wustl.edu/gsc
e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:

all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/OTB/CHK7> or send an E-mail to egreen@nhgri.nih.gov

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996). VECTOR: pBEO Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of H_RG016J04; actual end is at 156214 of H_RG016J04. The orientation of this clone is unknown.

This clone contains STS SWSS2784 (NID:g1113580) and SWSS893 (NID:g454733).

FEATURES

source	Location/Qualifiers
repeat_region	1..156214 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7q21" /clone="RG016J04" /clone.lib="CITB-978SK-B" 15..40 /rpt_family="L1"
repeat_region	326..621 /rpt_family="ALU"
repeat_region	complement(977..1499) /rpt_family="L1"
repeat_region	3398..3421 /rpt_family="L1"
repeat_region	5319..5345 /rpt_family="L1"
repeat_region	complement(8071..8347) /rpt_family="ALU"
repeat_region	complement(9406..9975) /rpt_family="L1"
repeat_region	complement(10000..11285) /rpt_family="L1"
repeat_region	complement(11315..11984) /rpt_family="L1"
repeat_region	11666..11687 /rpt_family="L1"
repeat_region	complement(12010..12299) /rpt_family="ALU"
repeat_region	complement(12301..13893) /rpt_family="L1"
repeat_region	13897..14096 /rpt_family="L1"
repeat_region	complement(14777..14838) /rpt_family="L1"
misc_feature	15715..15767 /note="similar to human EST T02878 (NID:g314119)"
misc_feature	19436..19497 /note="similar to human EST AA123941 (NID:g1682616)"
repeat_region	mq22f09.r1 complement(25113..25148) /rpt_family="L1"
repeat_region	complement(25561..25578)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:55:18 ; Search time 4356.39 seconds
(without alignments)
1272.965 Million cell updates/sec

Title: US-09-802-520-7
Perfect score: 265
Sequence: 1 ctgactctggtatctgcaa.....ggagggaagtaagccaagg 265

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----	-----	-----	-----	-----	-----

C	1	244	92.1	2453	6	AX155249	Sequence	7 from Patent WO0140276.	DNA	linear	PAT 22-JUN-2001
C	2	150	56.6	156214	9	HSAC002064	Human BAC				
	3	99	37.4	162928	2	AC104475	Pan trogl				
	4	93	35.1	250	6	AX106432	Sequence				
	5	93	35.1	250	6	AX140723	Sequence				
	6	93	35.1	250	6	AX200583	Sequence				
	7	93	35.1	250	6	AX267239	Sequence				
	8	86	32.5	366	6	AX106434	Sequence				
	9	86	32.5	366	6	AX140725	Sequence				
	10	86	32.5	366	6	AX200585	Sequence				
	11	86	32.5	366	6	AX267241	Sequence				
C	12	85	32.1	172915	2	AC099742	Papio cyn				
C	13	21	7.9	183	6	AX155254	Sequence				
	14	21	7.9	301	6	AX106507	Sequence				
	15	21	7.9	301	6	AX140798	Sequence				
	16	21	7.9	301	6	AX200658	Sequence				
	17	21	7.9	301	6	AX267314	Sequence				
C	18	21	7.9	507	6	AX341401	Sequence				
C	19	21	7.9	2412	9	AK026813	Homo sapi				
	20	21	7.9	63921	8	AB026639	Arabidops				
C	21	21	7.9	87401	2	AC021898	Homo sapi				
C	22	20	7.5	47065	2	AC102896	Mus muscu				
	23	20	7.5	69663	2	AC090747	Homo sapi				
C	24	20	7.5	73778	2	AC090743	Homo sapi				
	25	20	7.5	96000	9	AC079684	Homo sapi				
	26	20	7.5	97580	9	HS591N18	Human DNA				
	27	20	7.5	107806	9	AL445438	Human DNA				
	28	20	7.5	162990	2	AC024707	Homo sapi				
	29	20	7.5	163241	2	AL355472	Homo sapi				
	30	20	7.5	168498	2	AL583847	Homo sapi				
	31	20	7.5	169670	9	AC023908	Homo sapi				
C	32	20	7.5	172506	2	AC106467	Rattus no				
	33	20	7.5	174115	2	AC026138	Homo sapi				
	34	20	7.5	197782	9	AC012377	Homo sapi				
C	35	19	7.2	49870	9	AP003729	Homo sapi				
	36	19	7.2	57787	2	AC104880	Mus muscu				
	37	19	7.2	67529	2	AC069312	Homo sapi				
	38	19	7.2	85101	9	AL450244	Human DNA				
	39	19	7.2	96270	2	AC069369	Homo sapi				
C	40	19	7.2	99367	2	AC098899	Rattus no				
C	41	19	7.2	100168	2	AC110431	Rattus no				
C	42	19	7.2	106704	2	AC008574	Homo sapi				
C	43	19	7.2	112843	9	AP003165	Homo sapi				
C	44	19	7.2	148407	2	AC097603	Rattus no				
C	45	19	7.2	156912	2	AC016787	Homo sapi				

ALIGNMENTS

RESULT 1

AX155249/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

THIS PAGE BLANK (USPTO)

```

BE152388/c
LOCUS          BE152388          404 bp    mRNA    linear    EST 21-JUN-2000
DEFINITION     CM2-HT0323-171199-033-d11 HT0323 Homo sapiens cDNA, mRNA sequence.
ACCESSION      BE152388
VERSION        BE152388.1  GI:8615109
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS        Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
               B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.,
               and Fraser,C.M.
TITLE          Mouse BAC End Sequences from Library RPCI-23
JOURNAL        Other GSSs: RPCI-23-30713.TJ
MEDLINE        Unpublished (1999)
COMMENT        Contact: Shaying Zhao
               Department of Eukaryotic Genomics
               The Institute for Genomic Research
               9712 Medical Center Dr., Rockville, MD 20850, USA
               Tel: 301 838 0200
               Fax: 301 838 0208
               Email: szhao@tigr.org
               Clones are derived from the mouse BAC library RPCI-23. For BAC
               library availability, please contact Pieter de Jong
               (pieter@dejong.med.buffalo.edu). Clones may be purchased from
               BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
               or from Resea ch Genetics (info@resgen.com). BAC end page:
               http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
               Plate: 307 row: I column: 3
               Seq primer: T7
               Class: BAC ends.
               Location/Qualifiers
                 1..493
                   /organism="Mus musculus"
                   /strain="C57BL/6J"
                   /db_xref="taxon:10090"
                   /clone="RPCI-23-30713"
                   /clone_lib="RPCI-23"
                   /sex="Female"
                   /lab_host="DH10B"
                   /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
                   EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
                   brain genomic DNA was isolated and partially digested
                   with a combination of EcoRI and EcoRI Methyase. Size
                   selected DNA was cloned into the pBACe3.6 vector at the
                   EcoRI sites. The ligation products were transformed into
                   DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT    133 a 82 c 84 g 193 t
ORIGIN
Query Match          3.4%; Score 19; DB 12; Length 493;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 153 tcttctggcagctgcttat 171
      |||||
Db 301 TCTTCTGGCAGCTGCTTAT 283

Search completed: September 20, 2002, 08:33:51
Job time: 8715 sec

BE152388/c
LOCUS          BE152388          404 bp    mRNA    linear    EST 21-JUN-2000
DEFINITION     CM2-HT0323-171199-033-d11 HT0323 Homo sapiens cDNA, mRNA sequence.
ACCESSION      BE152388
VERSION        BE152388.1  GI:8615109
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS        Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
               Nagai,M.A., da Silva,W., Jr., Zago,M.A., Bordin,S., Costa,F.F.,
               Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
               Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
               M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
               Simpson,A.J.
TITLE          Shotgun sequencing of the human transcriptome with ORF expressed
               sequence tags
JOURNAL        Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE        20202663
COMMENT        Contact: Simpson A.J.G.
               Laboratory of Cancer Genetics
               Ludwig Institute for Cancer Research
               Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
               Brazil
               Tel: +55-11-2704922
               Fax: +55-11-2707001
               Email: asimpson@ludwig.org.br
               This sequence was derived from the FAPESP/LICR Human Cancer Genome
               Project. This entry can be seen in the following URL
               (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-CM2-HT0323-171
               199-033-d11&t3=1999-11-17&t4=1)
               Seq primer: puc 18 forward
               High quality sequence start: 7
               High quality sequence stop: 404.
               Location/Qualifiers
                 1..404
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone_lib="HT0323"
                   /dev_stage="Adult"
                   /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                   Site_2: SmaI; A mini-library was made by cloning products
                   derived from ORESTES PCR (U.S. Letters Patent application
                   No. 196,716 - Ludwig Institute for Cancer Research)
                   profiles into the pUC 18 vector. Reverse transcription of
                   tissue mRNA and cDNA amplification were performed under
                   low stringency conditions."
BASE COUNT    112 a 83 c 90 g 119 t
ORIGIN
Query Match          3.4%; Score 19; DB 9; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 tgttgctacagcctctgc 297
      |||||
Db 337 TGTTCCTACAGCCTCTGTC 319

RESULT 15
LOCUS          AQ983972/c
DEFINITION     RPCI-23-30713-TV RPCI-23 Mus musculus genomic clone RPCI-23-30713,
               DNA sequence.
ACCESSION      AQ983972
VERSION        AQ983972.1  GI:6817177
KEYWORDS       GSS.
SOURCE         house mouse.
ORGANISM       Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 880 row: D column: 8
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 488.

FEATURES

source
 Location/Qualifiers
 1..488
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-880 Col=8 Row=D"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
 BASE COUNT 175 a 85 c 73 g 153 t 2 others
 ORIGIN

Query Match 3.6%; Score 20; DB 12; Length 488;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 496 agagaattcagttttattca 515
 |||||
 Db 300 AGAGAATTCAAGTTTATTCA 281

RESULT 12

LOCUS 224497/c 306 bp mRNA linear EST 26-JUL-1993
 DEFINITION ATTS939 Ra147.1 Arabidopsis thaliana cDNA clone RaR062 5', mRNA sequence.
 ACCESSION 224497
 VERSION 224497.1 GI:394846
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 REFERENCE 1 (bases 1 to 306)
 AUTHORS CNRS.
 TITLE The Arabidopsis thaliana transcribed genome: the GDR cDNA program
 JOURNAL Unpublished (1996)
 COMMENT Contact: Lacomme,C. and Roby,D.
 UMR05 CNRS/INRA
 BP 27,31326 Castanet-Tolosan cedex,France.
 Email: roby@coulouse.inra.fr.
 Location/Qualifiers
 1..306

FEATURES

source
 Location/Qualifiers
 1..306
 /organism="Arabidopsis thaliana"
 /strain="ecotype Columbia"
 /db_xref="taxon:3702"
 /clone="RaR062"
 /clone_lib="Ra147.1"
 /note="vector: Lambda ZAPII; Physiological conditions:
 cycling cells one hour post-inoculation with X campestris
 pv campestris, strain 147 in the presence of cycloneximide
 1 micromolar."
 BASE COUNT 99 a 53 c 72 g 82 t

Query Match

3.4%; Score 19; DB 10; Length 306;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 472 tcagtgcagcaatgctttaa 490
 |||||
 Db 231 TCAGTGAGCAATGCTTTAA 213

RESULT 13

LOCUS T53018 377 bp mRNA linear EST 06-FEB-1995
 DEFINITION Ya82h06.s1 Stratagene ovary (#937217) Homo sapiens cDNA clone IMAGE:68219 3', mRNA sequence.

ACCESSION T53018
 VERSION T53018.1 GI:654878
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 377)

AUTHORS

Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478

TITLE

JOURNAL ya82h06.rl

MEDLINE

COMMENT

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

Insert Size: 21

High quality sequence stops: 345 Source: IMAGE Consortium, LNL This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -21ml3

High quality sequence stop: 345.

FEATURES

source

Location/Qualifiers
 1..377
 /organism="Homo sapiens"
 /db_xref="GDB:504476"
 /db_xref="taxon:9606"
 /clone="IMAGE:68219"
 /clone_lib="Stratagene ovary (#937217)"
 /sex="female"
 /dev_stage="49 year old"
 /lab_host="SOLR cells (kanamycin resistant)"
 /note="Organ: ovary; Vector: Bluescript SK; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt.
 Total ovary tissue, normal, caucasian. Average insert size: 0.8 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTT 3'"

BASE COUNT

92 a 82 c 79 g 124 t

ORIGIN

Query Match 3.4%; Score 19; DB 10; Length 377;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 gcaggtctctggcagctg 166
 |||||
 Db 118 GCAGGTCTCTGGCAGCTG 136

RESULT 14

DEFINITION PM2-AN0093-151000-003-a09 AN0093 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF733741
 VERSION BF733741.1 GI:12058977
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 261)
 Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 CONTACT: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM2&t2=PM2-AN0093-151000-003-a09&t3=2000-10-15&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 3
 High quality sequence stop: 261.
 High quality sequence stop: 261.
 Location/Qualifiers
 1..261
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="AN0093"
 /dev_stage="Adult"
 /note="Organ: amnion_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 81 a 53 c 40 g 87 t
 ORIGIN
 Query Match 3.6%; Score 20; DB 10; Length 261;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 216 ggaacctgtggtacagtga 235
 |||||||
 Db 63 GGAAACCTGGTTACAGTGA 44
 RESULT 10
 AA511666
 LOCUS
 DEFINITION vj29c03.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
 clone IMAGE:930436 5', mRNA sequence.
 ACCESSION AA511666
 VERSION AA511666.1 GI:2249520
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 380)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wyllie, F., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 The WashU-HHMI Mouse EST Project
 Unpublished (1996)
 CONTACT: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:535356
 Seq primer: -28ml3 rev1 ET from Amersham
 High quality sequence stop: 329.
 Location/Qualifiers
 1..380
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:930436"
 /clone_lib="Stratagene mouse diaphragm (#937303)"
 /tissue_type="diaphragm"
 /dev_stage="adult"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: diaphragm; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA prepared from diaphragm muscle. Primer: Oligo dt. Average insert size: 1.5 kb. Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGACGTTTTTTTTTTT 3"
 BASE COUNT 108 a 83 c 75 g 114 t
 ORIGIN
 Query Match 3.6%; Score 20; DB 9; Length 380;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 46 tatgctagaacacacagag 65
 |||||||
 Db 352 TATGCTAGAAACCAACAGAG 371
 RESULT 11
 AQ833520
 LOCUS
 DEFINITION HS_5304_B2_B04_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=880 Col=8 Row=D, DNA sequence.
 ACCESSION AQ833520
 VERSION AQ833520.1 GI:5799582
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 488)
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 99380589
 CONTACT: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887

AUTHORS
 TITLE
 JOURNAL
 COMMENT
 WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:535356
 Seq primer: -28ml3 rev1 ET from Amersham
 High quality sequence stop: 329.
 Location/Qualifiers
 1..380
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:930436"
 /clone_lib="Stratagene mouse diaphragm (#937303)"
 /tissue_type="diaphragm"
 /dev_stage="adult"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: diaphragm; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA prepared from diaphragm muscle. Primer: Oligo dt. Average insert size: 1.5 kb. Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGACGTTTTTTTTTTT 3"
 BASE COUNT 108 a 83 c 75 g 114 t
 ORIGIN
 Query Match 3.6%; Score 20; DB 9; Length 380;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 46 tatgctagaacacacagag 65
 |||||||
 Db 352 TATGCTAGAAACCAACAGAG 371
 RESULT 11
 AQ833520/c
 LOCUS
 DEFINITION HS_5304_B2_B04_T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=880 Col=8 Row=D, DNA sequence.
 ACCESSION AQ833520
 VERSION AQ833520.1 GI:5799582
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 488)
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 99380589
 CONTACT: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887

201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES

source
Location/Qualifiers
1. .728
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCAQC12"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 214 a 108 c 134 g 270 t 2 others
ORIGIN

Query Match 3.9%; Score 22; DB 9; Length 728;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 cttgatgcctttgtcagagatg 34
|||||
DB 197 CTTGATGCGCTTTGTGAGAGATG 218

RESULT

BH067591/c 7
LOCUS BH067591 627 bp DNA linear GSS 18-JUL-2001
DEFINITION RPCI-24-256N22.TV RPCI-24 Mus musculus genomic clone RPCI-24-256N22
ACCESSION BH067591.1 GI:14886972
VERSION BH067591.1
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS 1 (bases 1 to 627)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shwartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-24-256N22.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 256 Row: N Column: 22
Seq primer: 17
Class: BAC ends.

FEATURES

source
Location/Qualifiers
1. .627
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-256N22"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 180 a 126 c 130 g 191 t
ORIGIN

Query Match 3.8%; Score 21; DB 12; Length 627;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tttttttcttgatgccttt 24
|||||
DB 310 TTTTTCCTTGATGCGCTTT 290

RESULT

AQ544394 8
LOCUS AQ544394 198 bp DNA linear GSS 28-MAY-1999
DEFINITION CITBI-E1-2608017.TR CITBI-E1 Homo sapiens genomic clone 2608017, DNA sequence.
ACCESSION AQ544394
VERSION AQ544394.1 GI:4903469
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 198)
Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and Venter,J.C.
TITLE Use of BAC End Sequences from Caltech Libraries for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbsa@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES

source
Location/Qualifiers
1. .198
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2608017"
/clone_lib="CITBI-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"

BASE COUNT 73 a 25 c 39 g 61 t
ORIGIN

Query Match 3.6%; Score 20; DB 12; Length 198;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 481 aatgcttaaacctggagaga 500
|||||
DB 150 AATGCTTTAAACTGGAGAGA 169

RESULT

BH733741/c 9
LOCUS BH733741 261 bp mRNA linear EST 09-JAN-2001

```

Qy 382 gagagaagtttgagaattgaattatctctcttggcataatgagccttggtta 441
Db 514 GAGGAAGAAGTTTGAGAATTGAATGTATATCTCTTTGGCATAATGAGCCTTGCGCTTA 573

Qy 442 ctttcctcctggcagtcac 461
Db 574 CTTTCCCTCTGCGGAGTCAC 593

RESULT 4
BM431438 558 bp mRNA linear EST 31-JAN-2002
LOCUS lduo16f03 Bos taurus Duodenum #1 library Bos taurus CDNA, mRNA
DEFINITION
ACCESSION BM431438
VERSION BM431438.1 GI:18453160
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 558)
Hansen, C., Fu, A., Meng, Y., Li, C., Okine, E., Senses, C.W., Gordon
, P.M.K. and Moore, S.S.
Gene Expression Profiling of the Bovine Gastrointestinal Tract
Unpublished (2002)
Contact: Dr. Stephen Moore
. Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel: 780 492 0169
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca
Insert Length: 558 Std Error: 0.00
POLYA-No.
FEATURES
source
1. 558
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bos taurus Duodenum #1 library"
/tissue_type="Smooth muscle"
/cell_type="Simple columnar epithelial"
/dev_stage="Young adult"
/lab_host="X11-BlueMRF/strain"
/notes="Organ: Intestine/duodenum; Vector: Uni-2ZAPXR;
Site_1: EcoRI; Site_2: Xho I"
BASE COUNT 152 a 109 c 108 g 189 t
ORIGIN

Query Match 6.2%; Score 35; DB 10; Length 558;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 523 ctgtgatgtcgtctgctgctatgaagtactttcca 557
Db 58 CTTGGATATGCTGCTGCTGCTCATAAGTACTTTCCA 92

RESULT 5
AQ406733 497 bp DNA linear GSS 17-MAR-1999
LOCUS HS_5103_B2_A08_T7A RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate-679 Col-16 Row-B, DNA sequence.
ACCESSION AQ406733
VERSION AQ406733.1 GI:4429355
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 497)

```

```

AUTHORS Mahairas, G.C., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 679 row: B column: 16
Seq primer: T7
Class: BAC ends
High quality sequence stop: 497.
Location/Qualifiers
1. 497
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT 159 a 92 c 91 g 146 t
ORIGIN

Query Match 3.9%; Score 22; DB 12; Length 497;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 389 aagtttgagaattgaaatgta 410
Db 350 AAGTTGGAGAATTGAAATGTA 371

RESULT 6
AV646697 728 bp mRNA linear EST 15-JAN-2002
LOCUS AV646697 GLC Homo sapiens CDNA clone GLCAQC12 3', mRNA sequence.
DEFINITION
ACCESSION AV646697
VERSION AV646697.1 GI:9867711
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 728)
Xiao, H., Huang, J., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X.,
Shen, K., Lu, G., Fu, G., Zhong, M., Xu, S., Gu, W., Huang, W., Zhao, X.,
Hu, G., Gu, J., Chen, Z. and Han, Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
JOURNAL
MEDLINE
COMMENT Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-tech Park, Pudong, Shanghai

```

intraepithelial neoplasia (low-grade), cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."

BASE COUNT 84 a 70 C 62 g 106 t

ORIGIN

Query Match 31.1%; Score 174; DB 9; Length 322;

Best Local Similarity 99.6%; Pred. No. 1.3e-75; Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 35 tgattcattccatgctgtagaaccacagagtgactttttacaaaaatttcctatagagattg 94

|||||

DB 35 TGATTTCATCCCATATGCTAGAACCAACACAGAGTGACTTTTACAAAAATTCCTATAGAGATTG 94

|||||

QY 95 tgaataaaacctactactagtgccattacttctctccctagctatatacctcgaggctc 154

|||||

DB 95 TGAATAAAACCTTACTATAGTGGCATTACTTTTCTCTCCCTAGTATACCTTGCAGGTC 154

|||||

QY 155 ttctggcagctcttatcaactttattacggcaccaagtataggagatttccaccttgtt 214

|||||

DB 155 TTCTGGCAGCTCTTATCACTTTATTAGGCACCAAGTATAGGAGATTTCACACCTTGGT 214

|||||

QY 215 tgggaacctggttacagtgtagaaacacagcttggattactaagtt 259

|||||

DB 215 TGGAAACCTGGTTACAGTGTAGAAAACAGCTTGGATTACTAAGTT 259

|||||

RESULT 2

AI016724/c

LOCUS AI016724 393 bp mRNA linear EST 27-AUG-1998

DEFINITION OV03d03.x1 NCI_CGAP_K1d3 Homo sapiens cDNA clone IMAGE:1636229 3',

mRNA sequence.

ACCESSION AI016724

VERSION AI016724.1 GI:3231060

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 393)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

TITLE Unpublished (1997)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgapbs-re@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 305 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham.

Location/Qualifiers

1..393

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1636229"

/clone_lib="NCI_CGAP_K1d3"

/lab_host="DH10B"

/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer,

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not

I and Eco RI sites of the modified pT7T3 vector. mRNA

source: 2 pooled kidneys. Library went through one round

of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 132 a 84 c 61 g 116 t

ORIGIN

Query Match 22.7%; Score 127; DB 9; Length 393;

Best Local Similarity 100.0%; Pred. No. 3e-52; Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 agtttggagaattgaaatgtatatctccttggcataatgagccttgcttactttccct 449

|||||

DB 393 AGTTTGGAGAATTGAAATGTATATCTCCTTTGGCATAATGAGCCTTGGCTTACTTTCCCT 334

|||||

QY 450 cctggcagtcacttctatcccttcagtgagcaatgctttaaactggagagaattcagttt 509

|||||

DB 333 CCTGGCAGTCACCTTCTATCCCTTCAGTGACCAATGCTTTAAACTGGAGAAATTCAGTTT 274

|||||

QY 510 tatttcag 516

|||||

DB 273 TATTTCAG 267

|||||

RESULT 3

BI359597

LOCUS BI359597

DEFINITION 384188 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.

ACCESSION BI359597

VERSION BI359597.1 GI:15055625

KEYWORDS EST.

SOURCE pig.

ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 595)

AUTHORS Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,

Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.

and Keele, J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for

EST discovery in swine

JOURNAL Unpublished (2000)

COMMENT Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred

v0.980904 e. Vector identified by cross_match with the -minscore 18

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCCAGTCACGCG

Plate: 132 row: H column: 3

Seq primer: ATTTAGGTGACACTATAG.

Location/Qualifiers

1..595

/organism="Sus scrofa"

/db_xref="taxon:9823"

/clone_lib="MARC 2P1G"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;

Library made from pooled tissue from testis, ovary,

endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 143 a 138 c 126 g 188 t

ORIGIN

Query Match

Best Local Similarity 14.3%; Score 80; DB 10; Length 595;

Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 08:33:47 ; Search time 5053.42 Seconds
(without alignments)
1495.679 Million cell updates/sec

Title: US-09-802-520-6

Perfect score: 560

Sequence: 1 acatttttttcttgatgc.....ctcataagtaactttccatgt 560

Scoring table: OLIGO.NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:**
1: em_estba:**
2: em_esthum:**
3: em_estin:**
4: em_estnu:**
5: em_estov:**
6: em_estpl:**
7: em_estro:**
8: em_htc:**
9: gb_estl:**
10: gb_est2:**
11: gb_htc:**
12: gb_gss:**
13: em_gss_hum:**
14: em_gss_inv:**
15: em_gss_pln:**
16: em_gss_vrt:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	174	31.1	322	9	AA508880	AA508880 ng86f03.s
2	127	22.7	393	9	AI016724	AI016724 o703d03.x
3	80	14.3	595	10	BI359597	BI359597 384188 MA
4	35	6.2	558	10	BM431438	BM431438 lduo16f03
5	22	3.9	497	12	AQ406733	AQ406733 HS_5103_B
6	22	3.9	728	9	AV646697	AV646697 AV646697
7	21	3.8	627	12	BH067591	BH067591 RPCI-24-2
8	20	3.6	198	12	AQ544394	AQ544394 CITBT-EJ-
9	20	3.6	261	10	BF733741	BF733741 PM2-AN009
10	20	3.6	380	9	AA511666	AA511666 v129c03.r
11	20	3.6	488	12	AQ833520	AQ833520 HS_5304_B
12	19	3.4	306	10	Z24497	Z24497 ATTS939 Ra1
13	19	3.4	377	10	T53018	T53018 va82h06.s1
14	19	3.4	404	9	BE152388	BE152388 CM2-HT032
15	19	3.4	493	12	AQ983972	AQ983972 RPCI-23-3
16	19	3.4	505	10	BI350635	BI350635 fu34h06.y
17	19	3.4	560	10	BM034788	BM034788 fu35g08.y

18	19	3.4	561	10	BG625191	BG625191 pgnlc.pk0
19	19	3.4	561	10	BI133492	BI133492 UT-M-BH3-
20	19	3.4	563	10	BI845092	BI845092 fg35e10.y
21	19	3.4	563	10	BM035184	BM035184 fu40e11.y
22	19	3.4	572	12	AQ567312	AQ567312 HS_2110_B
23	19	3.4	577	10	BI864123	BI864123 ft15b04.y
24	19	3.4	578	9	AV933870	AV933870 AV933870
25	19	3.4	579	10	BI864118	BI864118 ft15a07.y
26	19	3.4	589	12	AZ748868	AZ748868 RPCI-24-8
27	19	3.4	614	10	BI472993	BI472993 rf94e01.y
28	19	3.4	622	12	AZ790856	AZ790856 2M0039K04
29	19	3.4	631	12	AZ939466	AZ939466 2M0198N23
30	19	3.4	636	10	BI839566	BI839566 fg39h04.y
31	19	3.4	644	10	BI428484	BI428484 ft84e12.y
32	19	3.4	659	9	AW305417	AW305417 fj59b06.y
33	19	3.4	672	12	AQ958620	AQ958620 LBRAY67TR
34	19	3.4	697	12	AZ897813	AZ897813 RPCI-24-2
35	19	3.4	720	10	BI704176	BI704176 zfshtARAM
36	19	3.4	729	10	BG622063	BG622063 602646669
37	19	3.4	777	12	CNSQ2M2K	AL204761 Tetraodon
38	19	3.4	796	12	BI0678	BI0678 F20D22-T7 I
39	19	3.4	1070	12	B08494	B08494 F20D22-T7-2
40	19	3.4	1102	10	BM465003	BM465003 AGENCOURT
41	19	3.4	1140	12	BI1198	BI1198 F20D22-T7.1
42	18	3.2	66	9	AA123941	AA123941 mq2fio9.r
43	18	3.2	113	9	AW902857	AW902857 QV3-NN102
44	18	3.2	196	9	AW690320	AW690320 NF029D05S
45	18	3.2	196	9	AA524308	AA524308 ng32h12.s

ALIGNMENTS

RESULT 1
AA508880
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

AA508880 ng86f03.s1 NCI_CGAP_Pr6 Homo sapiens cdna clone IMAGE:941693, mRNA sequence.
AA508880.1 GI:2245821
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 322)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuquai, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1000 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 249.
Location/Qualifiers
1..322
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:941693"
/clone_lib="NCI_CGAP_Pr6"
/sex="male"
/tissue_type="prostate"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from prostatic

THIS PAGE BLANK (USPTO)

Search completed: September 20, 2002, 09:51:16
Job time: 11864 sec

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2081
US-08-235-836C-71

Query Match 3.2%; Score 18; DB 4; Length 2081;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 162 agctgcttatcaacttta 179
|||||
Db 1064 AGCTGCTTATCAACTTTA 1047

RESULT 12
US-09-813-819-3/c
; Sequence 3, Application US/09813819
; Patent No. 6294368
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001177
; CURRENT APPLICATION NUMBER: US/09/813,819
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 17138
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(17138)
; OTHER INFORMATION: n = A,T,C or G
US-09-813-819-3

Query Match 3.2%; Score 18; DB 4; Length 17138;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 282 tgcctacagcctctgctt 299
|||||
Db 2862 TGCCTACAGCCTCTGCTT 2845

RESULT 13
US-09-920-048-3/c
; Sequence 3, Application US/09920048
; Patent No. 6344352
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001177DIV
; CURRENT APPLICATION NUMBER: US/09/920,048
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 09/813,819
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 17138
; TYPE: DNA
; ORGANISM: Human
; FEATURE:

; NAME/KEY: misc_feature
; LOCATION: (1)...(17138)
; OTHER INFORMATION: n = A,T,C or G
US-09-920-048-3

Query Match 3.2%; Score 18; DB 4; Length 17138;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 282 tgcctacagcctctgctt 299
|||||
Db 2862 TGCCTACAGCCTCTGCTT 2845

RESULT 14
US-08-943-731-141/c
; Sequence 141, Application US/08943731
; Patent No. 6265157
; GENERAL INFORMATION:
; APPLICANT: PROCKOP, DARWIN J.
; APPLICANT: SPOTILA, LORETTA D.
; APPLICANT: DELTAS, CONSTANTINOS D.
; APPLICANT: SEREDA, LARISA
; APPLICANT: LARSON, ANDREA W.
; APPLICANT: PACK, MICHAEL
; APPLICANT: COLIGE, ALAIN
; APPLICANT: EARLY, JAMES
; APPLICANT: KORKKO, JARMO
; APPLICANT: ALA-KOKKO, LEENA, et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
; NUMBER OF SEQUENCES: 666
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
; STREET: FLR.
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,731
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,322
; FILING DATE: 14-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/803,628
; FILING DATE: 03-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: DOYLE LEARY Ph.D., KATHRYN
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9598-27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-965-1284
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 141:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-943-731-141

Qy 91 attgtgaataaaaccttacctatagttggccattacttctctcccta 138
|||||
Db 388 ATTGTGAATAAACCTTACCTATAGTTGCCATTACTTTGCTCTCCCTA 435

RESULT 9
US-09-439-313-214
; Sequence 214, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 214
; LENGTH: 444
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(444)
; OTHER INFORMATION: n = A,T,C or G
US-09-439-313-214

Query Match 8.6%; Score 48; DB 4; Length 444;
Best Local Similarity 100.0%; Pred. No. 6.1e-15;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 attgtgaataaaaccttacctatagttggccattacttctctcccta 138
|||||
Db 388 attgtgaataaaaccttacctatagttggccattacttctctcccta 435

RESULT 10
US-08-235-836C-67/c
; Sequence 67, Application US/08235836C
; Patent No. 6248562
; GENERAL INFORMATION:
; APPLICANT: Dunn, John J.
; APPLICANT: Luft, Benjamin J.
; TITLE OF INVENTION: No. 6248562el Chimeric Proteins Comprising
; TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brookhaven National Laboratory
; STREET: Upton
; CITY: Upton
; STATE: NY
; COUNTRY: USA
; ZIP: 11973
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235,836C
; FILING DATE: 29-APR-1994

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,191
; FILING DATE: 01-11-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Bogosian, Margaret C.
; REGISTRATION NUMBER: 25,324
; REFERENCE/DOCKET NUMBER: BNL93-28A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 282-7338
; TELEFAX: (516) 282-3729
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2081 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2081
US-08-235-836C-67

Query Match 3.2%; Score 18; DB 4; Length 2081;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 162 agctgcttatacacttta 179
|||||
Db 1064 AGCTGCTTATCACTTTA 1047

RESULT 11
US-08-235-836C-71/c
; Sequence 71, Application US/08235836C
; Patent No. 6248562
; GENERAL INFORMATION:
; APPLICANT: Dunn, John J.
; APPLICANT: Luft, Benjamin J.
; TITLE OF INVENTION: No. 6248562el Chimeric Proteins Comprising
; TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brookhaven National Laboratory
; STREET: Upton
; CITY: Upton
; STATE: NY
; COUNTRY: USA
; ZIP: 11973
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235,836C
; FILING DATE: 29-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,191
; FILING DATE: 01-11-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Bogosian, Margaret C.
; REGISTRATION NUMBER: 25,324
; REFERENCE/DOCKET NUMBER: BNL93-28A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 282-7338
; TELEFAX: (516) 282-3729
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2081 base pairs
; TYPE: nucleic acid

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 213:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 250 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-030-607-213

Query Match 23.6%; Score 132; DB 4; Length 250;
Best Local Similarity 99.5%; Pred. No. 4.8e-57;
Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 342 ttatcagcaggttcataatgcaaaactcttggaatgaggaagtttgagaat 401
DB 208 ttatcagcaggttcataatgcaaaactcttggaatgaggaagtttgagaat 149
QY 402 tgaatgtatatctcttggcataatgagccttggttactttccctcctggcagtcac 461
DB 148 tgaatgtatatctcttggcataatgagccttggttactttccctcctggcagtcac 89
QY 462 ttctatccctcagtgagcaatgctttaaactggagagaatttcagtttattcagttctac 521
DB 88 ttctatccctcagtgagcaatgctttaaactggagagaatttcagtttattcagttctac 29
QY 522 act 524
DB 28 ACT 26

RESULT 7
US-09-439-313-213/c
; Sequence 213, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang toqui
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C3
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 213
; LENGTH: 250
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:

; NAME/KEY: misc_feature
; LOCATION: (1)...(250)
; OTHER INFORMATION: n = A,T,C or G
US-09-439-313-213

Query Match 23.6%; Score 132; DB 4; Length 250;
Best Local Similarity 99.5%; Pred. No. 4.8e-57;
Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 342 ttatcagcaggttcataatgcaaaactcttggaatgaggaagtttgagaat 401
DB 208 ttatcagcaggttcataatgcaaaactcttggaatgaggaagtttgagaat 149
QY 402 tgaatgtatatctcttggcataatgagccttggttactttccctcctggcagtcac 461
DB 148 tgaatgtatatctcttggcataatgagccttggttactttccctcctggcagtcac 89
QY 462 ttctatccctcagtgagcaatgctttaaactggagagaatttcagtttattcagttctac 521
DB 88 ttctatccctcagtgagcaatgctttaaactggagagaatttcagtttattcagttctac 29
QY 522 act 524
DB 28 ACT 26

RESULT 8
US-09-030-607-214
; Sequence 214, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 214:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 444 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-030-607-214

Query Match 8.6%; Score 48; DB 4; Length 444;
Best Local Similarity 100.0%; Pred. No. 6.1e-15;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db 285 TTCGCTATGGTCCATGTTGGCTACAGCCTCTGCTTACCGATGAGAAGGTCAGAGATAT 226
Qy 325 ttgtttctcaacatggcttatcagcaggttcatcaaatattgaaactcttgaatgag 384
Db 225 TTGTTCTCAACATGGCTTATCAGCAGGTTCTATGCAATATTGAAACTCTTGAATGAG 166
Qy 385 gaagaagtttgagaattgaaatgtatatctctcttggcgaataatgagccttggcttactt 444
Db 165 GAAGAAGTTTGAGAATTGAAATGTATATNTCCTTTGGCAATATCAGCCTTGGCTTACTT 106
Qy 445 tccctctggcagtcactctctatcccttcagtgagcaatgctttaaactggagagaattc 504
Db 105 TCCCTCTGGCAGTCACCTTCTATCCCTTCACTGAGCAATGCTTTAAACTGGAGAAATTC 46
Qy 505 agttttattcagt 517
Db 45 AGTTTATTTCAGT 33

RESULT 4
US-09-439-313-215/c
; Sequence 215, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqui
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 366
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(366)
; OTHER INFORMATION: n = A,T,C or G
US-09-439-313-215

Query Match 36.1%; Score 202; DB 4; Length 366;
Best Local Similarity 99.6%; Pred. No. 3.9e-92;
Matches 252; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 265 ttgcgtatgttccatgttgcttaacagcctctgttaccgatgagaaggtcagagagatat 324
Db 285 TTCGCTATGGTCCATGTTGGCTACAGCCTCTGCTTACCGATGAGAAGGTCAGAGATAT 226
Qy 325 ttgtttctcaacatggcttatcagcaggttcatcaaatattgaaactcttgaatgag 384
Db 225 TTGTTCTCAACATGGCTTATCAGCAGGTTCTATGCAATATTGAAACTCTTGAATGAG 166
Qy 385 gaagaagtttgagaattgaaatgtatatctctcttggcgaataatgagccttggcttactt 444
Db 165 GAAGAAGTTTGAGAATTGAAATGTATATNTCCTTTGGCAATATGAGCCTTGGCTTACTT 106
Qy 445 tccctctggcagtcactctctatcccttcagtgagcaatgctttaaactggagagaattc 504
Db 105 TCCCTCTGGCAGTCACCTTCTATCCCTTCACTGAGCAATGCTTTAAACTGGAGAAATTC 46

```

```

Qy 505 agttttattcagt 517
Db 45 AGTTTATTTCAGT 33

RESULT 5
US-09-323-873A-9
; Sequence 9, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; TITLE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF
; FILE REFERENCE: 129.16USU2
; CURRENT APPLICATION NUMBER: US/09/323,873A
; CURRENT FILING DATE: 1999-06-01
; PRIOR FILING DATE: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR FILING DATE: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 322
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-323-873A-9

Query Match 31.1%; Score 174; DB 4; Length 322;
Best Local Similarity 99.6%; Pred. No. 4.2e-78;
Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 35 tgattcatcatatgcttagaaccacacagagcttcttcaaaaattcctataagattg 94
Db 35 tgattcatcatatgcttagaaccacacagagcttcttcaaaaattcctataagattg 94
Qy 95 tgaataaaaccttaccctatagttgccattacttctctccctagtagtataccctcaggtc 154
Db 95 tgaataaaaccttaccctatagttgccattacttctctccctagtagtataccctcaggtc 154
Qy 155 ttctggcagctgttatacaactttattacggcaccacagtagagagatttccaccttgg 214
Db 155 ttctggcagctgttatacaactttattacggcaccacagtagagagatttccaccttgg 214
Qy 215 tggaaacctggttacagtgtagaacaacagcttgattactaagtt 259
Db 215 tggaaacctggttacagtgtagaacaacagcttgattactaagtt 259

RESULT 6
US-09-030-607-213/c
; Sequence 213, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

Db 241 gagagagtcagagagatatgtttctcaacatggcttatcagcaggttcatagcgcaaatat 300
Qy 366 tgaataactcttggaatgaggaagatttgagaaattgaaatgataatctctcttggcat 425
Db 301 tgaataactcttggaatgaggaagatttgagaaattgaaatgataatctctcttggcat 360
Qy 426 aatgagccttggtctacttccctcctcctgcagtcactctatccctcagtgagcaatgc 485
Db 361 aatgagccttggtctacttccctcctcctgcagtcactctatccctcagtgagcaatgc 420
Qy 486 tttaactggagagaatcagttttatcagtcactcacaacttgagatgctgctctcctcat 545
Db 421 tttaactggagagaatcagttttatcagtcactcacaacttgagatgctgctctcctcat 480
Qy 546 aagtactttccatgt 560
Db 481 aagtactttccatgt 495
RESULT 2
US-09-083-521-3
; Sequence 3, Application US/09083521
; Patent No. 6048970
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,521
; FILING DATE: Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0527 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1213 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT10
; CLONE: 1691243
US-09-083-521-3

Query Match 57.3%; Score 321; DB 3; Length 1213;
Best Local Similarity 99.7%; Pred. No. 8.7e-152;
Matches 37; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 189 caagtataggagattccaccttggttggaacactggttacagtgtagaacaacagcttgg 248
Db 1 CAAGTATAGGAGATTCCACCTTGGTGGAAACCTGGTTACAGTGTAGAAACACAGCTTGG 60

Qy 249 attactaagttttattcttcgctatggtccatggttcctcacagcctctgcttaccgatgag 308
Db 61 ATTACTAAGTTTTCCTTCGCTATGTTGCTCATGTTGCTTACAGCCTCTGCTTACCCGATGAG 120
Qy 309 aagtcagagagatatgtttctcaacatggcttatcagcaggttcacgaaatatiga 368
Db 121 AAGGTCAGAGAGATATGTTTCTCAACATGGCTTATCAGCAGGTTTCATGCAAAATATTGA 180
Qy 369 aaactcttgggaatgaggaagatttgagaaattgaaatgataatctctcttggcataat 428
Db 181 AAACCTCTTGGGAATGAGGAAGATTGGAGAAATTCAAATGTATATCTCTCTTTGGCATAAT 240
Qy 429 gagccttggtctacttccctcctcctgcagtcactctatccctcagtgagcaatgcttt 488
Db 241 GAGCCTTGGCTTACTTTCCTCCTGGCAGTCACCTTCTATCCCTTCAGTGAGCAATGCTTT 300
Qy 489 aaactggagagaatcagttttatcagtcactcacaacttgagatgctgctctcctcataag 548
Db 301 AAACCTGGAGAGANTTCAGTTTATTTCAGTCTACACTTGGATATGCTGCTCTGCTCATAG 360
Qy 549 tactttccatgt 560
Db 361 TACTTTCCATGT 372
RESULT 3
US-09-030-607-215/c
; Sequence 215, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 215:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 366 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-030-607-215

Query Match 36.1%; Score 202; DB 4; Length 366;
Best Local Similarity 99.6%; Pred. No. 3.9e-92;
Matches 252; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 265 ttcgtaggttcctcatgttgcctacagcctctgcttaccgatgagaggtcagagatat 324
Db 1 ttcgtaggttcctcatgttgcctacagcctctgcttaccgatgagaggtcagagatat 324

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:51:09 ; Search time 139.75 Seconds
(without alignments)
984.292 Million cell updates

Title: US-09-802-520-6

perfect score:

Sequence: 1 acatttttttccttgatgc.....ctcataagtactttccatgt 560

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0.

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB' seq length: 0

Maximum DB seq length: 200000000
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_patents_NA:*

```
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
```

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/2/ina/pctus_comb.seq:*

```
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	444	79.3	521	4	US-09-323-873A-7	Sequence 7, Appli
2	321	57.3	1213	3	US-09-083-521-3	Sequence 3, Appli
C 3	202	36.1	366	4	US-09-030-607-215	Sequence 215, App
C 4	202	36.1	366	4	US-09-439-313-215	Sequence 215, App
C 5	174	31.1	322	4	US-09-323-873A-9	Sequence 9, Appli
C 6	132	23.6	250	4	US-09-030-607-213	Sequence 213, App
C 7	132	23.6	250	4	US-09-439-313-213	Sequence 213, App
C 8	48	8.6	444	4	US-09-030-607-214	Sequence 214, App
9	48	8.6	444	4	US-09-439-313-214	Sequence 214, App
C 10	18	3.2	2081	4	US-08-235-836C-67	Sequence 67, Appli
C 11	18	3.2	2081	4	US-08-235-836C-71	Sequence 71, Appli
C 12	18	3.2	17138	4	US-09-813-819-3	Sequence 3, Appli
C 13	18	3.2	17138	4	US-09-920-048-3	Sequence 3, Appli
C 14	17	3.0	511	4	US-08-943-731-141	Sequence 141, App
15	17	3.0	1515	1	US-08-221-816B-1	Sequence 1, Appli
16	17	3.0	1687	1	US-08-143-219-26	Sequence 26, Appli
17	17	3.0	1975	2	US-08-852-743-1	Sequence 1, Appli
18	17	3.0	1975	3	US-09-185-370-1	Sequence 1, Appli
19	17	3.0	2161	2	US-08-712-709-4	Sequence 4, Appli
20	17	3.0	2161	3	US-09-111-444-4	Sequence 4, Appli
21	17	3.0	2161	4	US-09-541-228-4	Sequence 4, Appli
C 22	17	3.0	24183	4	US-08-943-731-3	Sequence 3, Appli
23	16	2.9	558	4	US-09-328-111-84	Sequence 84, Appli
24	16	2.9	602	1	US-08-682-218-16	Sequence 16, Appli
25	16	2.9	605	1	US-08-682-218-11	Sequence 11, Appli
26	16	2.9	605	1	US-08-682-218-12	Sequence 12, Appli
27	16	2.9	606	1	US-08-682-218-13	Sequence 13, Appli

ALIGNMENTS

```

RESULT 1
US-09-323-873A-7
; Sequence 7, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Safran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; FILE REFERENCE: 129.160SU2
; CURRENT APPLICATION NUMBER: US/09/323.873A
; CURRENT FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-323-873A-7

```

	Query Match	79.3%;	Score 444;	DB 4;	Length 521;	
	Best Local Similarity	99.8%;	Pred. NO. 1.9e-213;			
	Matches 494; Conservative	0;	Mismatches 1;	Indels	0;	Gaps 0;
<hr/>						
Qy	66	tgacttttacaaaattcctatagagatttggaaataaaccttaccatagtgcattac	125			
Dd	1	tgacttttacaaaattcctatagagatttggaaataaaccttaccatagtgcattac	60			
<hr/>						
Qy	126	tttgcttccttagtatacctcgaggtctctctggcagctgcttacaactttattacgg	185			
Dd	61	tttgcttccttagtatacctcgaggtctctctggcagctgcttacaactttattacgg	120			
<hr/>						
Qy	186	caccaagtataggagattccaccttggttggaaaacctggttacagtgtagaaaaacagct	245			
Dd	121	caccaagtataggagattccaccttggttggaaaacctggttacagtgtagaaaaacagct	180			
<hr/>						
Qy	246	tggattactaagtttttattcttcgtcta tgg tccatgttgtcctacaagcctctgcttaccgat	305			
Dd	181	tggattactaagttttttcttcgtcta tgg tccatgttgtcctacaagcctctgcttaccgat	240			
<hr/>						
Qy	306	qadaaqgtcaagaqagatatgtttctccaacttgcttatacagcaggtttcatgcaaatac	365			

THIS PAGE BLANK (USPTO)

XX The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 395 BP; 100 A; 84 C; 80 G; 131 T; 0 other;
 Query Match 50.4%; Score 282; DB 22; Length 395;
 Best Local Similarity 99.7%; Pred. No. 4.2e-135;
 Matches 332; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 20 ccttgcagagatgtgattccatccatgctagaaacccaacagagtgactttacaaaa 79
 DB 63 ccttgcagagatgtgattccatccatgctagaaacccaacagagtgactttacaaaa 122
 QY 80 ttccatagagattgtaataaaaccttaccatgcttagtgccattcttgcctccctag 139
 DB 123 ttccatagagattgtaataaaaccttaccatgcttagtgccattcttgcctccctag 182
 QY 140 tatacctgcaggtctctcggcagctcttaccatgcttagtgccattcttgcctccctag 199
 DB 183 tatacctgcaggtctctcggcagctcttaccatgcttagtgccattcttgcctccctag 242
 QY 200 gattccacacctgttggaacacctgttgcagtgtagaataaacagcttggattactaagt 259
 DB 243 gattccacacctgttggaacacctgttgcagtgtagaataaacagcttggattactaagt 302
 QY 260 ttattctgcgtatggtccatgttgcctacagcctctgcttaccgatgagaaggtcagaga 319
 DB 303 ttattctgcgtatggtccatgttgcctacagcctctgcttaccgatgagaaggtcagaga 362
 QY 320 gataattgtttctcaacatggcttatcagcagg 352
 DB 363 gataattgtttctcaacatggcttatcagcagg 395
 RESULT 15
 AAS15797
 ID AAS15797 standard; DNA; 528 BP.
 XX
 AC AAS15797;
 XX
 DT 16-JAN-2002 (first entry)
 XX
 DE Human Six-Transmembrane Protein of Prostate 1, STMP1, exon 4.
 XX
 KW Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
 KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
 KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
 KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
 KW liver cancer; lung cancer; cytostatic; ds; exon 4.
 XX
 OS Homo sapiens.
 XX
 PN WO200172962-A2.
 XX
 PD 04-OCT-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09410.
 XX
 PR 24-MAR-2000; 2000US-191929P.
 XX
 PA (SAAT/) SAATCIOGLU F.
 XX
 PI Saaticoglu F;
 XX
 DR WPI; 2001-662926/76.
 XX
 PT New polynucleotide for the diagnosis, prevention and treatment for
 PT prostate and testis disorders, particularly prostate cancer, comprises

PT prostate-specific or testis-specific nucleic acids
 XX
 PS Claim 6; Fig 4C; 114pp; English.
 XX
 CC The invention relates to substantially pure prostate-specific or
 CC testis-specific polypeptides and the nucleic acids encoding them.
 CC Also included are vectors and host cells expressing the proteins, a
 CC transgenic animal expressing the protein, antibodies against the
 CC proteins, probes for detecting the nucleic acids, antisense molecules
 CC for the nucleic acids and methods of isolating modulators of the
 CC proteins. Compounds that modulate the prostate specific or testis
 CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,
 CC cryptorchidism, undescended, retractile, ascending or vanished
 CC testis. Other proliferative disorders for which the modulators may be
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The
 CC present sequence represents exon 4 of a prostate specific protein,
 CC Six-Transmembrane Protein of Prostate 1, STMP1.
 XX
 SQ Sequence 528 BP; 139 A; 116 C; 102 G; 171 T; 0 other;
 Query Match 50.2%; Score 281; DB 22; Length 528;
 Best Local Similarity 99.7%; Pred. No. 1.4e-134;
 Matches 331; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 20 ccttgcagagatgtgattccatccatgctagaaacccaacagagtgactttacaaaa 79
 DB 197 ccttgcagagatgtgattccatccatgctagaaacccaacagagtgactttacaaaa 256
 QY 80 ttccatagagattgtaataaaaccttaccatgcttagtgccattcttgcctccctag 139
 DB 257 ttccatagagattgtaataaaaccttaccatgcttagtgccattcttgcctccctag 316
 QY 140 tatacctgcaggtctctcggcagctcttaccatgcttagtgccattcttgcctccctag 199
 DB 317 tatacctgcaggtctctcggcagctcttaccatgcttagtgccattcttgcctccctag 376
 QY 200 gattccacacctgttggaacacctgttgcagtgtagaataaacagcttggattactaagt 259
 DB 377 gattccacacctgttggaacacctgttgcagtgtagaataaacagcttggattactaagt 436
 QY 260 ttattctgcgtatggtccatgttgcctacagcctctgcttaccgatgagaaggtcagaga 319
 DB 437 ttattctgcgtatggtccatgttgcctacagcctctgcttaccgatgagaaggtcagaga 496
 QY 320 gataattgtttctcaacatggcttatcagcagg 351
 DB 497 gataattgtttctcaacatggcttatcagcagg 528
 Search completed: September 20, 2002, 09:59:54
 Job time: 11712 sec

```
Db 63 cctttgcagagatggtgattccatccatgctagatgagaacccaacagagtgacttttcaaaa 122
QY 80 ttctctagagattggaataaaaccttacctatagttgccattactttgtctcctag 139
Db 123 ttctctagagattggaataaaaccttacctatagttgccattactttgtctcctag 182
QY 140 tatactcgcaggtctctctggcagctgcttatcaactttattacggcaccagaatagga 199
Db 183 tatactcgcaggtctctctggcagctgcttatcaactttattacggcaccagaatagga 242
QY 200 gatttccacctgtgttggaacctgtttacagtgtagaataacacagcttgattactaagt 259
Db 243 gatttccacctgtgttggaacctgtttacagtgtagaataacacagcttgattactaagt 302
QY 260 ttattctcgtatggtccatgttgctctacagcctctgcttaccgatgagaaggtcagaga 319
Db 303 ttctctcgtatggtccatgttgctctacagcctctgcttaccgatgagaaggtcagaga 362
QY 320 gatattgtttctcaacatggcttatcagcagg 352
Db 363 gatattgtttctcaacatggcttatcagcagg 395

RESULT 13
AAI15504
ID AAI15504 standard; DNA; 395 BP.
AC AAI15504;
XX
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #5437 for gene expression analysis in human cervical cell sample.
DE Probe: human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID No 5437; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENP). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
```

```
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 395 BP; 100 A; 84 C; 80 G; 131 T; 0 other;
XX
XX Query Match 50.4%; Score 282; DB 22; Length 395;
XX Best Local Similarity 99.7%; Pred. No. 4.2e-135;
XX Matches 332; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 20 cctttgcagagatggtgattccatccatgctagatgagaacccaacagagtgacttttcaaaa 79
Db 63 cctttgcagagatggtgattccatccatgctagatgagaacccaacagagtgacttttcaaaa 122
QY 80 ttctctagagattggaataaaaccttacctatagttgccattactttgtctcctag 139
Db 123 ttctctagagattggaataaaaccttacctatagttgccattactttgtctcctag 182
QY 140 tatactcgcaggtctctctggcagctgcttatcaactttattacggcaccagaatagga 199
Db 183 tatactcgcaggtctctctggcagctgcttatcaactttattacggcaccagaatagga 242
QY 200 gatttccacctgtgttggaacctgtttacagtgtagaataacacagcttgattactaagt 259
Db 243 gatttccacctgtgttggaacctgtttacagtgtagaataacacagcttgattactaagt 302
QY 260 ttattctcgtatggtccatgttgctctacagcctctgcttaccgatgagaaggtcagaga 319
Db 303 ttctctcgtatggtccatgttgctctacagcctctgcttaccgatgagaaggtcagaga 362
QY 320 gatattgtttctcaacatggcttatcagcagg 352
Db 363 gatattgtttctcaacatggcttatcagcagg 395

RESULT 14
AAI36922
ID AAI36922 standard; DNA; 395 BP.
XX
XX AC AAI36922;
XX
XX 17-OCT-2001 (first entry)
XX
XX DE Probe #5608 used to measure gene expression in human placenta sample.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157272-A2.
XX
XX PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488997/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 5608; 654pp; English.
```

Qy	20	cccttgcagagatgtgattcattcccatatgctagaacccaacagagtgcacttttaca	79
Db	63	cccttgcagagatgtgattcattcccatatgctagaacccaacagagtgcacttttaca	122
Qy	80	ttcctatagagatgtgaaataaaacccctaacctatgtgcattacttgcctccctag	139
Db	123	ttcctatagagatgtgaaataaaacccctaacctatgtgcattacttgcctccctag	182

```
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human fetal liver.
XX PS Claim 1; SEQ ID NO 5682; 639pp + sequence listing; English.
XX CC The invention relates to a single exon nucleic acid probe for
XX CC measuring human gene expression in a sample derived from human foetal
XX CC liver. The single exon nucleic acid probes may be used for predicting,
XX CC measuring and displaying gene expression in samples derived from human
XX CC fetal liver. The present sequence is a single exon nucleic acid
XX CC probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 395 BP; 100 A; 84 C; 80 G; 131 T; 0 other;

Query Match 50.4%; Score 282; DB 22; Length 395;
Best Local Similarity 99.7%; Pred. No. 4.2e-135;
Matches 332; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 20 ccttgcagagatgtgattccatcatgctagaaccacacagagtgactttacaaaa 79
DB |||||||
DB 63 ccttgcagagatgtgattccatcatgctagaaccacacagagtgactttacaaaa 122
QY 80 ttccatagagatgtgataaaaccttacctatagttgccattacttgcctccctag 139
DB |||||||
DB 123 ttccatagagatgtgataaaaccttacctatagttgccattacttgcctccctag 182
QY 140 tatactgcaggtctcttgcagctgtctatcaattattacgcaccagatagga 199
DB |||||||
DB 183 tatactgcaggtctcttgcagctgtctatcaattattacgcaccagatagga 242
QY 200 gattccacccttggttggaacctgttaccagttacagttagaaaaacagcttgattactaagt 259
DB |||||||
DB 243 gattccacccttggttggaacctgttaccagttacagttagaaaaacagcttgattactaagt 302
QY 260 ttattctgcctatggtccatgcttgcacagcctctgtctaccgatgagaaggtcagaga 319
DB |||||||
DB 303 tttcttcgctatggtccatggttgcctacagcctctgtctaccgatgagaaggtcagaga 362
QY 320 gatatttttttccaaatgcttatccagcagg 352
DB |||||||
DB 363 gatatttttttccaaatgcttatccagcagg 395

RESULT 10
ID ABA26909 standard; DNA; 395 BP.
```

```

Best Local Similarity 99.6%; Pred. No. 3.8e-171;
Matches 452; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 20 ccttgcagagatgattccatccatgtagaaccacacagagcttttcaaaa 79
Db 859 ccttgcagagatgattccatccatgtagaaccacacagagcttttcaaaa 918
Qy 80 ttctctagagatgtagaataaaccttaacctatagtttgccattacttgcctccctag 139
Db 919 ttctctagagatgtagaataaaccttaacctatagtttgccattacttgcctccctag 978
Qy 140 tataacctgcaggtcttctgcagctgcttatcaactttattacggcaccacagtatagga 199
Db 979 tataacctgcaggtcttctgcagctgcttatcaactttattacggcaccacagtatagga 1038
Qy 200 gattccacacttggttggaacctggttacagttgtagaataaacagcttggtattactagtt 259
Db 1039 gattccacacttggttggaacctggttacagttgtagaataaacagcttggtattactagtt 1098
Qy 260 ttattctgcctatggtccatgttgcctacagcctctgcttaccgatgagaaggtcagaga 319
Db 1099 ttattctgcctatggtccatgttgcctacagcctctgcttaccgatgagaaggtcagaga 1158
Qy 320 gatattgtttctcaacatggtcttatcagcaggttctatgcaaatattgaaaactcttggga 379
Db 1159 gatattgtttctcaacatggtcttatcagcaggttctatgcaaatattgaaaactcttggga 1218
Qy 380 atgaggagaagatttgagaattgaaatgatatctctcttggcataatgagccttggtct 439
Db 1219 atgaggagaagatttgagaattgaaatgatatctctcttggcataatgagccttggtct 1278
Qy 440 tactttccctctgagcagctacttctatcccttc 473
Db 1279 tactttccctctgagcagctacttctatcccttc 1312

RESULT 8
AAZ46296
ID AAZ46296 standard; cDNA; 1213 BP.
XX
AC AAZ46296;
XX
XX 07-MAR-2000 (first entry)
XX Human prostate growth-associated membrane protein PGAMP-1 cDNA..
XX Prostate growth-associated membrane protein; PGAMP-1; prostate;
XX consensus; antibody; screening; modulator; agonist; antagonist;
XX therapeutic agent; cancer; solid tumour; leukaemia; lymphoma;
XX reproductive disorder; infertility; endometriosis;
XX polycystic ovarian syndrome; prostatitis; recombinant expression;
XX gene therapy; antisense therapy; ribozyme; diagnosis; diagnosis;
XX monitoring; immunoassay; targeting; drug delivery; drug screening; ds.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 83..508
XX /*tag= a
XX /product= "Human PGAMP-1"
XX
XX WO961469-A2.
XX
XX 02-DEC-1999.
XX
XX 17-MAY-1999; 99WO-US10888.
XX
XX 22-MAY-1998; 98US-0083521.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Lal P, Guegler KJ, Corley NC;
XX

```

WPI; 2000-062671/05.
P-PSDB; AAY52589.

New human prostate growth-associated membrane proteins, for treating or preventing cancer and reproductive disorders

Claim 7; Page 67; 72pp; English.

This sequence represents cDNA encoding human prostate growth-associated protein PGAMP-1. Nucleotides encoding PGAMP-1 were initially identified in a prostate cDNA library, this sequence representing a consensus. Human prostate growth associated membrane proteins PGAMP-1 and PGAMP-2 (AAY52590) may be used to raise specific antibodies and to screen for specific modulators (agonists, antagonists or other potential therapeutic agents). Antagonists of PGAMP are used to treat or prevent a wide range of cancers (solid tumours, leukaemia, lymphoma etc.) and reproductive disorders (such as infertility, endometriosis, polycystic ovarian syndrome, prostatitis). PGAMP-encoding nucleic acids, its fragments and complements, may be used for recombinant production of PGAMP proteins, in gene therapy (e.g., as antisense molecules, triplex-forming molecules and ribozymes), and as diagnostic probes and primers. Anti-PGAMP antibodies may be used for diagnosis and monitoring of PGAMP-related diseases by standard immunoassays, as therapeutic antagonists (including targeted delivery of other drugs), and in competitive drug screens.

Sequence 1213 BP; 335 A; 239 C; 215 G; 424 T; 0 other;

Query Match 57.3%; Score 321; DB 21; Length 1213;
Best Local Similarity 99.7%; Pred. No. 3.5e-155;
Matches 371; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 189 caagtataggagatttccacacttggttggaacacctggttacagttgtagaacaacagcttgg 248
Db 1 caagtataggagatttccacacttggttggaacacctggttacagttgtagaacaacagcttgg 60
Qy 249 attactaagttttatcttcgctatggtccatggttgcctacagcctctgcttaccgtagag 308
Db 61 attactaagttttatcttcgctatggtccatggttgcctacagcctctgcttaccgtagag 120
Qy 309 aaggtcagagagatattgtttctcaacatggtcttatcagcaggttccatgcaaatattga 368
Db 121 aaggtcagagagatattgtttctcaacatggtcttatcagcaggttccatgcaaatattga 180
Qy 369 aaactcttggaatgaggagaagtttggaagaattggaatgataatctctcttggcataat 428
Db 181 aaactcttggaatgaggagaagtttggaagaattggaatgataatctctcttggcataat 240
Qy 429 gagccttggcttactttccctccctggtcagtcacttctctcctcctcagtgagcaatgcttt 488
Db 241 gagccttggcttactttccctccctggtcagtcacttctctcctcctcagtgagcaatgcttt 300
Qy 489 aaactgagagaattcagttttatttcagttctacacttggttatgctcgtctgctcataag 548
Db 301 aaactgagagaattcagttttatttcagttctacacttggttatgctcgtctgctcataag 360
Qy 549 tactttccatgt 560
Db 361 tactttccatgt 372

RESULT 9
ABA57377
ID ABA57377 standard; DNA; 395 BP.
XX
XX ABA57377;
XX
XX 01-FEB-2002 (first entry)
XX Human foetal liver single exon nucleic acid probe #5682.
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.


```
QY 187 accaagtagagagatttccacacttgggtgaaacctgggttacagtgtagaataacagctt 246
Db 121 accaagtagagagatttccacacttgggtgaaacctgggttacagtgtagaataacagctt 180
QY 247 ggattactaaagtttactctgctatggtccatgtgctacagctctgctaccgatg 306
Db 181 ggattactaaagtttctctgctatggtccatgtgctacagctctgctaccgatg 240
QY 307 agaagtcagagagatattgtttctcaacatgggttatcagcaggttccatcaaatatt 366
Db 241 agaagtcagagagatattgtttctcaacatgggttatcagcaggttccatcaaatatt 300
QY 367 gaaactcttgaaagagaaagtttggaagtgaatgaaatgtatatctctctggcata 426
Db 301 gaaactcttgaaagagaaagtttggaagtgaatgaaatgtatatctctctggcata 360
QY 427 atgagccttgcttacttctcctcctggcagtcacactctctcctcctcagtgagcaatgt 486
Db 361 atgagccttgcttacttctcctcctggcagtcacactctctcctcctcagtgagcaatgt 420
QY 487 ttaactggagagaattcagttttattcagtcacacttggaatggtatgctgctgcata 546
Db 421 ttaactggagagaattcagttttattcagtcacacttggaatggtatgctgctgcata 480
QY 547 agtacttccatgt 560
Db 481 agtacttccatgt 494

RESULT 4
AAD07072
ID AAD07072 standard; cDNA; 2453 BP.
XX
AC AAD07072;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human six transmembrane epithelial antigen of prostate-2 clone GTD3 cDNA.
XX
KW Human; cytostatic; antiproliferative; vaccine; gene therapy;
KW six transmembrane epithelial antigen of the prostate-2; STEAP-2;
KW chromosome 7q21; cancer; prostate; colon; bladder; lung; ovarian;
KW pancreatic; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
RBS 352..358
FT /*tag= a
FT /*note= "Kozak region"
FT 355..1719
FT /*tag= b
FT /*product= "Human six transmembrane epithelial antigen
FT of the prostate (STEAP)-2"
FT 709..2073
FT /*tag= c
FT /*product= "Human six transmembrane epithelial antigen
FT of the prostate (STEAP)-2, alternative version"
FT /*note= "CDS does not include start and stop codon"
FT /*transl_except= (pos:1714..1722, aa:Asp-Ala)
FT /*transl_except= (pos:1834..1842, aa:Arg-Ser)
FT /*transl_except= (pos:1957..1965, aa:Glu-Gly)
FT /*transl_except= (pos:2050..2058, aa:Thr-Ser)
FT /*transl_except= (pos:2062..2070, aa:Asn-Phe)
FT /*note= "Inframe stop codon alters the reading frame"
FT /*partial
XX
PN W0200140276-A2.
XX
PD 07-JUN-2001.
XX
PF 06-DEC-2000; 2000WO-US33040.
XX
```

```
PR 06-DEC-1999; 99US-0455486.
PA (UROC-) UROGENESYS INC.
PI Afar DEH, Hubert RS, Raitano AB, Saffran DC, Mitchell SC, Faris M;
PI Jakobovits A;
XX
DR WPI; 2001-367804/38.
DR P-PSDB; AAE02781, AAE02841.
XX
PT New STEAP (six transmembrane epithelial antigen of the prostate)
PT proteins, expressed in human cancers, useful for detecting and treating
PT cancer -
XX
PS Claim 4; Fig 9A-9D; 187pp; English.
XX
CC The present sequence is human six transmembrane epithelial antigen of
CC the prostate (STEAP)-2 clone GTD3 cDNA. STEAP is a member of cell
CC surface serpentine transmembrane antigens. STEAP-2 gene is located on
CC chromosome 7q21 and is used in gene therapy. Inhibiting the development
CC or progression of a cancer (eg. prostate, colon, bladder, lung, ovarian
CC and pancreatic) expressing STEAP or inhibiting growth or killing cells
CC expressing STEAP in a patient, comprises administering a vaccine
CC composition to the patient. Treating a patient with a cancer that
CC compresses STEAP, or inhibiting growth or killing cells expressing STEAP,
CC comprises administering to the patient a vector encoding single chain
CC monoclonal antibody that comprises the variable domains of the heavy and
CC light chains of the monoclonal antibody that specifically binds to STEAP,
CC such that the vector delivers the single chain monoclonal antibody coding
CC sequence to the cancer cells and the encoded single chain monoclonal
CC antibody is expressed intracellularly.
CC Note: The present sequence is also shown in sequence listing of the
CC specification, but it lacks nucleotides at its 5' end.
XX
SQ Sequence 2453 BP; 648 A; 537 C; 520 G; 748 T; 0 other;

Query Match 78.4%; Score 439; DB 22; Length 2453;
Best Local Similarity 99.6%; Pred. No. 6.le-216;
Matches 539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 20 ccttctcagagatgtgattccatcatgtctagaaacacacagagtgacttttcaaaa 79
Db 1043 ccttctcagagatgtgattccatcatgtctagaaacacacagagtgacttttcaaaa 1102
QY 80 ttccctatagagattgtgaataaaaccttaccattagtgtccattacttctgtccctag 139
Db 1103 ttccctatagagattgtgaataaaaccttaccattagtgtccattacttctgtccctag 1162
QY 140 tatacctcgcaggttctctgagcagtgcttatcaactttattacggcaccagtatagga 199
Db 1163 tatacctcgcaggttctctgagcagtgcttatcaactttattacggcaccagtatagga 1222
QY 200 gatttccaccttgggtgaaacctgttaccagtttagaaaacacagcttgattactaagt 259
Db 1223 gatttccaccttgggtgaaacctgttaccagtttagaaaacacagcttgattactaagt 1282
QY 260 ttattctgcctatggtccatggttgcctacagcctctgtctaccgatgagaaggtcagaga 319
Db 1283 ttattctgcctatggtccatggttgcctacagcctctgtctaccgatgagaaggtcagaga 1342
QY 320 gatatttcttctcaacatggttattacagcaggttccatgcaaatattgaaaactcttga 379
Db 1343 gatatttcttctcaacatggttattacagcaggttccatgcaaatattgaaaactcttga 1402
QY 380 atgaggaagaagtttgggaattgaaatgtatatctcttctggcataatgagccttggct 439
Db 1403 atgaggaagaagtttgggaattgaaatgtatatctcttctggcataatgagccttggct 1462
QY 440 tactttccctcctcctgcagtcactctctatccctcagtcagtcagtcactgttaaaactggag 499
Db 1463 tactttccctcctcctgcagtcactctctatccctcagtcagtcagtcactgttaaaactggag 1522
```


XX PI Saatcioglu F;
XX WPI; 2001-662926/76.
DR P-PSDB; AAU10188.
XX
XX New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids -
XX
XX Claim 4; Fig 4H; 114pp; English.
XX
XX The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence encodes prostate specific protein, Six-Transmembrane
CC Protein of Prostate 1, STMP1, ORF2.
XX
XX Sequence 1561 BP; 420 A; 319 C; 346 G; 476 T; 0 other;

Query Match 87.5%; Score 490; DB 22; Length 1561;
Best Local Similarity 99.8%; Pred. No. 3.3e-242;
Matches 540; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 20 cctttgcagagatgtgattccatcatgctagataacacacagagtgacttttacaataa 79
Db 876 cctttgcagagatgtgattccatcatgctagataacacacagagtgacttttacaataa 935
QY 80 ttccatagagatttgataaaaccttacctatagtgccattacttgcctccctag 139
Db 936 ttccatagagatttgataaaaccttacctatagtgccattacttgcctccctag 995
QY 140 tatacctgcaggttcttgcagctgcttatacctaacttattacggcaccagtaggata 199
Db 996 tatacctgcaggttcttgcagctgcttatacctaacttattacggcaccagtaggata 1055
QY 200 gatttccaccttggtgaaacctggttacagtgtagaataacacagcttgattactaagt 259
Db 1056 gatttccaccttggtgaaacctggttacagtgtagaataacacagcttgattactaagt 1115
QY 260 ttatttcctatggtccatgtgctacagcctgcttaccgagtagaagggtcagaga 319
Db 1116 ttatttcctatggtccatgtgctacagcctgcttaccgagtagaagggtcagaga 1175
QY 320 gatatttgcctcaacatggtctatcatcagcagggttcctgcaaatattgaaaactcttga 379
Db 1176 gatatttgcctcaacatggtctatcatcagcagggttcctgcaaatattgaaaactcttga 1235
QY 380 atgaggaagaagtttggaataatgataatctctcttggcataatgagccttgct 439
Db 1236 atgaggaagaagtttggaataatgataatctctcttggcataatgagccttgct 1295
QY 440 tacttccctcctgcagctcattctatcccttcagtgagcaatgctttaaactggagag 499
Db 1296 tacttccctcctgcagctcattctatcccttcagtgagcaatgctttaaactggagag 1355
QY 500 aatcaggttttattcagttacacttgatgattgctgtcgtcgtcgtcgtcgtcgtcgtcgt 559
Db 1356 aatcaggttttattcagttacacttgatgattgctgtcgtcgtcgtcgtcgtcgtcgtcgt 1415
QY 560 t 560

Db 1416 t 1416
RESULT 2
AAS15810
ID AAS15810 standard; cDNA; 2238 BP.
XX AAS15810;
AC AAS15810;
XX
DT 16-JAN-2002 (first entry)
XX
DE Human ORF2 of Six-Transmembrane Protein of Prostate 1, STMP1.
XX
KW Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytotstatic; ss; ORF2.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 188..1552
FT /*tag= a
FT /product= "STMP1, ORF2"
XX
PN WO200172962-A2.
XX
PD 04-OCT-2001.
XX
PF 23-MAR-2001; 2001WO-US09410.
XX
PR 24-MAR-2000; 2000US-191929P.
XX
PA (SAAT/) SAATCIOGLU F.
XX
PI Saatcioglu F;
XX
DR WPI; 2001-662926/76.
XX P-PSDB; AAU10188.
XX
PT New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids -
XX
PS Claim 5; Fig 4G; 114pp; English.
XX
CC The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence represents the second open reading frame of a prostate
CC specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.
XX
SQ Sequence 2238 BP; 607 A; 457 C; 453 G; 721 T; 0 other;

Query Match 87.5%; Score 490; DB 22; Length 2238;
Best Local Similarity 99.8%; Pred. No. 3.3e-242;
Matches 540; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 20 cctttgcagagatgtgattccatcatgctagataacacacagagtgacttttacaataa 79

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:59:50 ; Search time 600.75 seconds
(without alignments)
1600.453 Million cell updates/sec

Title: US-09-802-520-6

Perfect score: 560

Sequence: 1 acattttttcttgatgc.....ctcataagtaacttccatgt 560

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	490	87.5	1561	22 AAS15811	Human cdna encodin
2	490	87.5	2238	22 AAS15810	Human ORF2 of Six-
3	443	79.1	519	21 AAD49398	Human STRAP-2 part
4	439	78.4	2453	22 AAD07072	Human six transmem
5	388	69.3	1680	22 AAS15802	Human cdna encodin
6	388	69.3	4329	22 AAS15801	Human ORF of Six-T
7	352	62.9	2102	22 AAS15813	Human cdna encodin
8	321	57.3	1213	21 AAD46296	Human prostate gro
9	282	50.4	395	22 ABA57377	Human foetal liver

10	282	50.4	395	22	ABA26909	Probe #5375 for ge
11	282	50.4	395	22	AAK05412	Human brain expres
12	282	50.4	395	22	AAK31011	Human bone marrow
13	282	50.4	395	22	AAI15504	Probe #5437 for ge
14	282	50.4	395	22	AAI36922	Probe #5608 used t
15	281	50.2	528	22	AAS15797	Human Six-Transmem
16	281	50.2	528	22	AAS15806	Human Six-Transmem
17	281	50.2	2192	23	AAS76493	DNA encoding novel
18	281	50.2	3900	23	AAS64300	DNA encoding novel
19	232	41.4	1725	22	AAS15793	Human DNA for Six-
c 20	202	36.1	366	19	AAV61349	Reverse DNA sequen
c 21	202	36.1	366	19	AAV58685	Prostate tumour sp
c 22	202	36.1	366	21	AAA06448	Human immunogenic
c 23	202	36.1	366	22	AAS63656	Human prostate cDN
c 24	202	36.1	366	22	AAS10207	Human prostate tum
c 25	202	36.1	366	22	AAH93584	Human prostate-spe
c 26	202	36.1	366	22	AAH84878	Human prostate-spe
c 27	202	36.1	366	22	AAH02629	Prostate tumour an
c 28	174	31.1	322	21	AAZ49399	Human STRAP-2 gene
c 29	174	31.1	322	22	AAD07074	NCI_CGAP Pr-6 cDNA
c 30	165	29.5	165	22	AAS15798	Human Six-Transmem
c 31	165	29.5	165	22	AAS15807	Human Six-Transmem
c 32	132	23.6	250	19	AAV61347	Reverse DNA sequen
c 33	132	23.6	250	19	AAV58683	Prostate tumour sp
c 34	132	23.6	250	21	AAA06446	Human immunogenic
c 35	132	23.6	250	22	AAS63654	Human prostate cDN
c 36	132	23.6	250	22	AAS10205	Human prostate tum
c 37	132	23.6	250	22	AAH93582	Human prostate-spe
c 38	132	23.6	250	22	AAH84876	Human prostate-spe
c 39	132	23.6	250	22	AAH02627	Prostate tumour an
c 40	89	15.9	116	22	ABA70004	Human foetal liver
c 41	89	15.9	116	22	ABA36829	Probe #15295 for g
c 42	89	15.9	116	22	AAK18212	Human brain expres
c 43	89	15.9	116	22	AAK44108	Human bone marrow
c 44	89	15.9	116	22	AAI24731	Probe #14664 for g
c 45	89	15.9	116	22	AAI50117	Probe #18803 used

ALIGNMENTS

RESULT 1

AAS15811

ID AAS15811 standard; cDNA; 1561 BP.

XX AC AAS15811;

XX AC AAS15811;

XX DT 16-JAN-2002 (first entry)

XX DT 16-JAN-2002 (first entry)

XX DE Human cdna encoding ORF2 of Six-Transmembrane Protein of Prostate 1.

XX DE Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ss; ORF2.

XX OS Homo sapiens.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 188..1552

XX FT /*tag= a

XX FT /product= "STMP1, ORF2"

XX PN WO200172962-A2.

XX PD 04-OCT-2001.

XX PF 23-MAR-2001; 2001WO-US09410.

XX PR 24-MAR-2000; 2000US-19129P.

XX PA (SAAT/) SAATCIOGLU F.

Query Match 9.5%; Score 53; DB 2; Length 87401;
Best Local Similarity 100.0%; Pred. No. 3.5e-18;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 cctttgtcagagatgctattcatcatctgtagaacaacacagagtgacttt 72
|||||
Db 46621 cctttgtcagagatgctattcatcatctgtagaacaacacagagtgacttt 46673
|||||

RESULT 15
AX106433
LOCUS AX106433 444 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 214 from Patent WO0125272.
ACCESSION AX106433
VERSION AX106433.1 GI:13922112
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 444)
AUTHORS Xu, J., Skeiky, Y.A., Reed, S.G. and Cheever, M.A.
TITLE Compositions and methods for therapy and diagnosis of prostate cancer
JOURNAL Patent: WO 0125272-A 214 12-APR-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..444
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..444
misc_feature /note="n = A,T,C or G"
BASE COUNT 120 a 100 c 78 g 143 t 3 others
ORIGIN

Query Match 8.6%; Score 48; DB 6; Length 444;
Best Local Similarity 100.0%; Pred. No. 2.7e-15;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 atttgtataaaaccttacctattgctccctta 138
|||||
Db 388 ATTGGAATAAAACCTTACCTATAGTGGCAATCTTGTCTCCCTA 435
|||||

* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1
* 902 1001: gap of 100 bp in length
* 1002 1907: contig of 906 bp in length
* 1908 2007: gap of 100 bp
* 2008 2912: contig of 905 bp in length
* 2913 3012: gap of 100 bp
* 3013 3872: contig of 860 bp in length
* 3873 3972: gap of 100 bp
* 3973 4868: contig of 896 bp in length
* 4869 4968: gap of 100 bp
* 4969 5874: contig of 906 bp in length
* 5875 5974: gap of 100 bp
* 5975 6897: contig of 923 bp in length
* 6898 6997: gap of 100 bp
* 6998 7903: contig of 906 bp in length
* 7904 8003: gap of 100 bp
* 8004 8913: contig of 910 bp in length
* 8914 9013: gap of 100 bp
* 9014 9906: contig of 893 bp in length
* 9907 10006: gap of 100 bp
* 10007 10906: contig of 900 bp in length
* 10907 11006: gap of 100 bp
* 11007 11910: contig of 904 bp in length
* 11911 12010: gap of 100 bp
* 12011 12924: contig of 914 bp in length
* 12925 13024: gap of 100 bp
* 13025 13930: contig of 906 bp in length
* 13931 14030: gap of 100 bp
* 14031 14919: contig of 889 bp in length
* 14920 15019: gap of 100 bp
* 15020 15900: contig of 881 bp in length
* 15901 16000: gap of 100 bp
* 16001 16938: contig of 938 bp in length
* 16939 17038: gap of 100 bp
* 17039 17953: contig of 915 bp in length
* 17954 18053: gap of 100 bp
* 18054 18908: contig of 855 bp in length
* 18909 19008: gap of 100 bp
* 19009 19906: contig of 898 bp in length
* 19907 20006: gap of 100 bp
* 20007 20922: contig of 916 bp in length
* 20923 21022: gap of 100 bp
* 21023 21936: contig of 914 bp in length
* 21937 22036: gap of 100 bp
* 22037 22942: contig of 906 bp in length
* 22943 23042: gap of 100 bp
* 23043 23928: contig of 886 bp in length
* 23929 24028: gap of 100 bp
* 24029 24911: contig of 883 bp in length
* 24912 25011: gap of 100 bp
* 25012 25887: contig of 876 bp in length
* 25888 25987: gap of 100 bp
* 25988 26883: contig of 896 bp in length
* 26884 26983: gap of 100 bp
* 26984 27912: contig of 929 bp in length
* 27913 28012: gap of 100 bp
* 28013 28919: contig of 907 bp in length
* 28920 29019: gap of 100 bp
* 29020 29933: contig of 914 bp in length
* 29934 30033: gap of 100 bp
* 30034 30930: contig of 897 bp in length
* 30931 31030: gap of 100 bp
* 31031 31924: contig of 894 bp in length
* 31925 32024: gap of 100 bp
* 32025 32902: contig of 878 bp in length
* 32903 33002: gap of 100 bp
* 33003 33890: contig of 888 bp in length
* 33891 33990: gap of 100 bp

* 33991 34900: contig of 910 bp in length
* 34901 35000: gap of 100 bp
* 35001 35895: contig of 895 bp in length
* 35896 35995: gap of 100 bp
* 35996 36840: contig of 845 bp in length
* 36841 36940: gap of 100 bp
* 36941 37822: contig of 882 bp in length
* 37823 37922: gap of 100 bp
* 37923 38845: contig of 923 bp in length
* 38846 38945: gap of 100 bp
* 38946 39823: contig of 878 bp in length
* 39824 39923: gap of 100 bp
* 39924 40792: contig of 869 bp in length
* 40793 40892: gap of 100 bp
* 40893 41811: contig of 919 bp in length
* 41812 41911: gap of 100 bp
* 41912 42806: contig of 895 bp in length
* 42807 42906: gap of 100 bp
* 42907 43784: contig of 878 bp in length
* 43785 43884: gap of 100 bp
* 43885 44777: contig of 893 bp in length
* 44778 44877: gap of 100 bp
* 44878 45768: contig of 891 bp in length
* 45769 45868: gap of 100 bp
* 45869 46761: contig of 893 bp in length
* 46762 46861: gap of 100 bp
* 46862 47713: contig of 852 bp in length
* 47714 47813: gap of 100 bp
* 47814 48729: contig of 916 bp in length
* 48730 48829: gap of 100 bp
* 48830 49779: contig of 950 bp in length
* 49780 49879: gap of 100 bp
* 49880 50794: contig of 915 bp in length
* 50795 50894: gap of 100 bp
* 50895 51797: contig of 903 bp in length
* 51798 51897: gap of 100 bp
* 51898 52786: contig of 889 bp in length
* 52787 52886: gap of 100 bp
* 52887 53779: contig of 893 bp in length
* 53780 53879: gap of 100 bp
* 53880 54782: contig of 903 bp in length
* 54783 54882: gap of 100 bp
* 54883 55769: contig of 887 bp in length
* 55770 55869: gap of 100 bp
* 55870 56736: contig of 867 bp in length
* 56737 56836: gap of 100 bp
* 56837 57715: contig of 879 bp in length
* 57716 57815: gap of 100 bp
* 57816 58712: contig of 897 bp in length
* 58713 58812: gap of 100 bp
* 58813 59729: contig of 917 bp in length
* 59730 59829: gap of 100 bp
* 59830 60638: contig of 809 bp in length
* 60639 60738: gap of 100 bp
* 60739 61654: contig of 916 bp in length
* 61655 61754: gap of 100 bp
* 61755 62680: contig of 926 bp in length
* 62681 62780: gap of 100 bp
* 62781 63664: contig of 884 bp in length
* 63665 63764: gap of 100 bp
* 63765 64670: contig of 906 bp in length
* 64671 64770: gap of 100 bp
* 64771 65666: contig of 896 bp in length
* 65667 65766: gap of 100 bp
* 65767 66636: contig of 870 bp in length
* 66637 66736: gap of 100 bp
* 66737 67604: contig of 868 bp in length
* 67605 67704: gap of 100 bp
* 67705 68576: contig of 872 bp in length
* 68577 68676: gap of 100 bp
* 68677 69601: contig of 925 bp in length
* 69602 69701: gap of 100 bp
* 69702 70618: contig of 917 bp in length

```

SOURCE      human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 250)
            Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
            Reed,S.G., Kalos,M.D., Fanger,G.R., Retter,M.W., Stolk,J.A.,
            Stolk,J.A., Skeiky,Y.A., Wang,A. and Meagher,M.J.
TITLE        Compositions and methods for the therapy and diagnosis of prostate
JOURNAL      cancer
PATENT:      WO 0151633-A 213 19-JUL-2001;
CORIXA       CORIXA CORPORATION (US)
FEATURES     Location/Qualifiers
            1..250
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT   81 a 53 c 42 g 69 t 5 others
ORIGIN
Query Match 23.6%; Score 132; DB 6; Length 250;
Best Local Similarity 99.5%; Pred. No. 3.8e-62;
Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 342 ttatcagcaggttcataatgaaactcttggaaatgaggaggaagtgttgagaat 401
Db 208 TTATCAGCAGGTTTCATGCAAAATATTGAAACTCTTGGAAATGAGGAAGATTGGAGAA 149

Qy 402 tgaatgatctcttggcataatgagcttgcttacttccctcctggcagtcac 461
Db 148 TGAATGATATNTCTTGGCATATAGAGCCTTGCTTACCTTCCCTCCGCGCAGTCAC 89

Qy 462 ttctatccctcagtgagcaatgctttaaactggagagaattcagttatttattcagtcac 521
Db 88 TTCTATCCCTTCAGTGAGCAATGCTTTAAACTGGAGAGAAATTCAGTTTATTTCAGTCTAC 29

Qy 522 act 524
Db 28 ACT 26

RESULT 14
AC021898 87401 bp DNA linear HTG 13-JUL-2000
LOCUS      Homo sapiens chromosome 15 clone RP11-407J8 map 15, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
ACCESSION AC021898
VERSION    AC021898.2 GI:9119282
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 87401)
            Birren,B., Linton,L., Nusbaum,C. and Lander,E.
            Homo sapiens chromosome 15, clone RP11-407J8
            Unpublished
            2 (bases 1 to 87401)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
            Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
            Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Jul 13, 2000 this sequence version replaced gi:6730901.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L2445
            Center clone name: 407_J_8
            -----
            * NOTE: This record contains 88 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows

```

	*	8134	8233: gap of unknown length	
	*	8234	40378: contig of 32145 bp in length	
	*	40379	40478: gap of unknown length	
	*	40479	172915: contig of 132437 bp in length.	
FEATURES			Location/Qualifiers	
source			1..172915	
			/organism="Papio cynocephalus anubis"	
			/db_xref="taxon:9555"	
			/clone="RP41-167P22"	
			/clone_lib="RP41"	
misc_feature			1..2438	
			/note="assembly_fragment clone_end:SP6 vector_side::right"	
misc_feature			2539..8133	
			/note="assembly_fragment"	
misc_feature			8234..40378	
			/note="assembly_fragment clone_end:T7 vector_side:right"	
misc_feature			40479..172915	
			/note="assembly_fragment"	
BASE COUNT		52509 a	31733 c 32277 g 56096 t	300 others
ORIGIN				
			Query Match	29.8%; Score 167; DB 2; Length 172915;
			Best Local Similarity	100.0%; Pred. No. 8.6e-82;
			Matches 167; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy		350	agggtcatgcaaatattgaaaactcttggaatggagggagaagtgttgagaaattgaaatgt 409	
Db		88840	AGGTTTCATGC AAAATTGGAAAACCTCTTGG AATGAGGAAGAAGTTTGGA GAATTCGAAATGT 88899	
Qy		410	atatcctcttggcataatgagcccttgcttaactttcccctctgcagtcaccttatcc 469	
Db		88900	ATA TCTCTTTGGCAT AATGAGCCTTGCTTACTTTCCCTCGCAGTCACTTCTATCC 88959	
Qy		470	cttcagtgagcaatgcctttaactggagagaattcagttttattcacg 516	
Db		88960	CTTCAGTGAGCAATGCTTTAAACTGGAGAGAAATTCAGTTTTATTTCAG 89006	
RESULT 10				
AXI06432/c				
LOCUS		AXI06432	250 bp DNA linear	PAT 30-APR-2001
DEFINITION		Sequence 213 from Patent WO0125272.		
ACCESSION		AXI06432		
VERSION		AXI06432.1 GI:13922111		
KEYWORDS		human.		
SOURCE				
ORGANISM		Homo sapiens		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE		1 (bases 1 to 250)		
AUTHORS		Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.		
TITLE		Compositions and methods for therapy and diagnosis of prostate cancer		
JOURNAL		Patent: WO 0125272-A 213 12-APR-2001;		
FEATURES		CORIXA CORPORATION (US)		
source		Location/Qualifiers		
		1..250		
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
misc_feature		1..250		
		/note="n = A,T,C or G"		
BASE COUNT		81 a	53 c 42 g 69 t	5 others
ORIGIN				
			Query Match	23.6%; Score 132; DB 6; Length 250;
			Best Local Similarity	99.5%; Pred. No. 3.8e-62;
			Matches 182; Conservative	0; Mismatches 1; Indels 0; Gaps 0;


```
/note="assembly_fragment"
66293. .92168
/note="assembly_fragment"
92269. .122036
/note="assembly_fragment"
122137. .162928
/note="assembly_fragment"
vector_side:right"
BASE COUNT 51528 a 29878 c 30533 g 50485 t 504 others
ORIGIN

Query Match 34.1%; Score 191; DB 2; Length 162928;
Best Local Similarity 99.6%; Pred. No. 3.5e-95;
Matches 241; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 20 ccttgtcagagatgattcattcatatgcttagaacaacacagagtgactttacaaa 79
|||||
Db 121604 CTTTGTGACAGATGATTCATCATATGCTAGAACCAACAGAGTGACTTTTACAAA 121545
|||||

QY 80 ttctcatagagattgtgaataaaacctaccattagttgccattactttgtctccctag 139
|||||
Db 121544 TTCCATATAGAGATTGATATAAACCTTACCTATAGTTCCTCTCTAG 121485
|||||

QY 140 tatactgcaggtctcttcggcagctgcttatacatttatacggcaccagtagtaga 199
|||||
Db 121484 TATACCTCGCAGGCTCTTCGGCAGCTGCTTATCACTTTATTACGGCACCAAGTAGGA 121425
|||||

QY 200 gatttccacctggttgaaacctggttacagtgtagaacaacagcttgattactaagt 259
|||||
Db 121424 GATTTCACCTGGTTGGAAACCTGTTACAGTGTAGAAAACAGCTTGGATTACTAAGTT 121365
|||||

QY 260 tt 261
||
Db 121364 TT 121363

RESULT 8
AX155253
LOCUS AX155253 322 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 11 from Patent WO0140276.
ACCESSION AX155253
VERSION AX155253.1 GI:14536717
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 322)
AUTHORS Afar,D.E., Hubert,R.S., Raitano,A.B., Saffran,D.C., Mitchell,S.C.,
Faris,M. and Jakobovits,A.
TITLE Serpentine transmembrane antigens expressed in human prostate
cancers and uses thereof
JOURNAL Patent: WO 0140276-A 11 07-JUN-2001;
Urogenesys, Inc. (US)
FEATURES
source
1..322
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 84 a 70 c 62 g 106 t
ORIGIN

Query Match 31.1%; Score 174; DB 6; Length 322;
Best Local Similarity 99.6%; Pred. No. 1.4e-85;
Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 35 tgattcattcatatgtagaaacacagagtgacttttacaataattctctatagattg 94
|||||
Db 35 TGATTTCATCCATATGCTAGAAACCAACAGAGTGACTTTTACAAAATTCCTATAGAGATTG 94
|||||

QY 95 tgaataaaacctaccattagttgcttacttctctccctagtagatataccttcgagggtc 154
|||||
Db 95 TGAATAAAACCTTACCTATAGTTGCCATTACTTGTCTCCCTAGTATACCTTGCAGGTC 154
|||||

QY 155 ttctggcagctgcttattcaactttattacggcaccagtagtaggatttccaccttgg 214
|||||
Db 155 TTCTGGCAGCTGCTTATCAACTTTATTACGGCACCAAGTATAGGAGATTTCACCTTGGT 214
|||||

QY 215 tggaaacctggttaccagtgtagaacaacagcttggtgattactaagtt 259
|||||
Db 215 TGGAAACCTGGTTACAGTGTAGAAAACAGCTTGGATTACTAAGTT 259
|||||

RESULT 9
AC099742
LOCUS AC099742 172915 bp DNA linear HTG 20-NOV-2001
DEFINITION Papio cynocephalus anubis clone RP41-167P22, WORKING DRAFT
SEQUENCE, 4 unordered pieces.
ACCESSION AC099742
VERSION AC099742.1 GI:17017546
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE olive baboon.
ORGANISM Papio cynocephalus anubis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Papio.
REFERENCE 1 (bases 1 to 172915)
AUTHORS Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Mastrian,S.D.,
McCluskey,J.C., McDowell,J., Pearson,R., Prasad,A., Shevchenko,K.,
Stantrapop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Zhang,L.-H. and Green,E.D.
TITLE NISC Comparative Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 172915)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (20-NOV-2001) NIH Intramural Sequencing Center, 8717
Grovesmont Circle, Gaithersburg, MD 20877, USA
COMMENT
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nhgri.nih.gov
----- Project Information
Center project name: ccy
Center clone name: 167P22
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171178 bases at least Q40
Consensus quality: 171910 bases at least Q30
Consensus quality: 172236 bases at least Q20
Insert size: 130000; agarose-fp
Quality coverage: 12.58x in Q20 bases; agarose-fp
Quality coverage: 9.47x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2438: contig of 2438 bp in length
* 2439 2538: gap of unknown length
* 2539 8133: contig of 5595 bp in length
* *
```

```

Db 165 GAAGAAGTTGGAGAAATGAAATGTATATNTCCCTTGGCATAATGAGCCTTGCGTTACTT 106
Qy 445 tccctccctggcagtcactctctatcccttcagtgagcaatgctttaaaactgagagaattc 504
Db 105 TCCCTCCGCGAGTCACCTCTCTATCCCTTCAGTGAGCAATGCTTTAAACTGGAGAGAAATC 46
Qy 505 agttttatttcagt 517
Db 45 AGTTTATTTCAGT 33

RESULT 6
LOCUS AX267241/c 366 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 215 from Patent WO0173032.
ACCESSION AX267241
VERSION AX267241.1 GI:16516019
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
Kalos, M.D., Fanger, G.R., Retter, M.W., Stolk, J.A., Day, C.H.,
Vedvick, T.S., Carter, D., Li, S.X., Wang, A., Skeiky, Y.A., Hepler, W.T.
and Henderson, R.A.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0173032-A 215 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
Source 1. 366
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 119 a 79 c 69 g 96 t 3 others
ORIGIN

Query Match 36.1%; Score 202; DB 6; Length 366;
Best Local Similarity 99.6%; Pred. No. 3.4e-101;
Matches 252; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 265 ttcgctatggtccatgttgctcagacgctctgcttacgcagatgagaaggtcagagatat 324
Db 285 TTCGCTATGGTCCATGTTGGCTACAGCCTCTGCTACCGATGAGAGGTGACAGAGATAT 226
Qy 325 ttgtttctcaacatggcttatcagcaggttctatgcaaatattgaaaactcttggaatgag 384
Db 225 TTGTTTCTCAACATGGCTTATCAGCAGGTTTCATGCAAAATFTGAAAACCTCTTGGAAATGAG 166
Qy 385 gaagaagtgtgagaattgaaatgatctatctctttggcataatgagcccttggttactt 444
Db 165 GAAGAAGTTGGAGAAATGAAATGTATATNTCCCTTGGCATAATGAGCCTTGCGTTACTT 106
Qy 445 tccctccctggcagtcactctctatcccttcagtgagcaatgctttaaaactgagagaattc 504
Db 105 TCCCTCCGCGAGTCACCTCTCTATCCCTTCAGTGAGCAATGCTTTAAACTGGAGAGAAATC 46
Qy 505 agttttatttcagt 517
Db 45 AGTTTATTTCAGT 33

RESULT 7
AC104475/c
LOCUS AC104475
DEFINITION Pan troglodytes clone RP43-120K11, WORKING DRAFT SEQUENCE, 6
unordered pieces.
ACCESSION AC104475
VERSION AC104475.1 GI:17530717
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```

```

SOURCE chimpanzee.
ORGANISM Pan troglodytes
Mammalia; Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 162928)
AUTHORS Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-O.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Masiello, C., Mastrian, S.D.,
McCloskey, J.C., Mcbowell, J., Pearson, R., Prasad, A., Stantirpop, S.,
Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L.,
Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and
Green, E.D.
TITLE NISC Comparative Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 162928)
AUTHORS Green, E.D.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2001) NTH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nhgri.nih.gov
----- Project Information
Center project name: ces
Center clone name: 120K11
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 161001 bases at least Q40
Consensus quality: 161422 bases at least Q30
Consensus quality: 161562 bases at least Q20
Insert size: 143000; agarose-fp
Quality coverage: 10.76x in Q20 bases; agarose-fp
Quality coverage: 9.47x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 13922: contig of 13922 bp in length
* 13923 14022: gap of unknown length
* 14023 36248: contig of 22226 bp in length
* 36249 36348: gap of unknown length
* 36349 66192: contig of 29844 bp in length
* 66193 66292: gap of unknown length
* 66293 92168: contig of 25876 bp in length
* 92169 92268: gap of unknown length
* 92269 122037: contig of 29768 bp in length
* 122037 122136: gap of unknown length
* 122137 162928: contig of 40792 bp in length.
FEATURES
Source
Location/Qualifiers
1. 162928
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-120K11"
/clone_lib="RP43"
1. 13922
/note="assembly_fragment"
clone_end:T7
vector_side:left"
14023..36248
/note="assembly_fragment"
36349..66192

```

```
RESULT 3
AX106434/c
LOCUS AX106434
DEFINITION Sequence 215 from Patent WO0125272.
ACCESSION AX106434
VERSION AX106434.1 GI:13922113
KEYWORDS human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 366)
AUTHORS Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.
TITLE Compositions and methods for therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0125272-A 215 12-APR-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1. 366
/organism="Homo sapiens"
/db_xref="taxon:9606"
misc_feature 119 a 79 c 69 g 96 t 3 others
BASE COUNT 119 a 79 c 69 g 96 t 3 others
ORIGIN
Query Match 36.1%; Score 202; DB 6; Length 366;
Best Local Similarity 99.6%; Pred. No. 3.4e-101;
Matches 252; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 265 ttccgtatggtccatgttgcttacagcctctgctaccgatgagaaggtcagagagatat 324
Db 285 TTTCGTATGGTCCATGTTGGCTTACAGCCTCTGCTTACCGATGAGAAGGTTCAGAGAGATAT 226
Qy 325 ttgtttctcaacatggtttatcagcaggttcatgcataattgaaactcttgggaatgag 384
Db 225 TTGTTTCTCAACATGGCTTATCAGCAGGTTTCATGCAATATTTGAAACTCTTGGGAATGAG 166
Qy 385 gaagaagtttgagaattgaaatgtatctctcttgccataatgagccttggcttactt 444
Db 165 GAAGAAGTTTGAGAAATGAAATGTATATNTCCCTTTGGCATAATGAGCCTTGGCTTACTT 106
Qy 445 tccctctggcagtcacttctatcccttcagtcagtcagtcagtcagtcagtcagtcagtc 504
Db 105 TCCCTCTGGCAGTCACCTTCTATCCCTTCAGTGAGCAATGCTTTAACTGGAGAGAAATTC 46
Qy 505 agttttatcagt 517
Db 45 AGTTTATTCAGT 33

Query Match 36.1%; Score 202; DB 6; Length 366;
Best Local Similarity 99.6%; Pred. No. 3.4e-101;
Matches 252; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 265 ttccgtatggtccatgttgcttacagcctctgctaccgatgagaaggtcagagagatat 324
Db 285 TTTCGTATGGTCCATGTTGGCTTACAGCCTCTGCTTACCGATGAGAAGGTTCAGAGAGATAT 226
Qy 325 ttgtttctcaacatggtttatcagcaggttcatgcataattgaaactcttgggaatgag 384
Db 225 TTGTTTCTCAACATGGCTTATCAGCAGGTTTCATGCAATATTTGAAACTCTTGGGAATGAG 166
Qy 385 gaagaagtttgagaattgaaatgtatctctcttgccataatgagccttggcttactt 444
Db 165 GAAGAAGTTTGAGAAATGAAATGTATATNTCCCTTTGGCATAATGAGCCTTGGCTTACTT 106
Qy 445 tccctctggcagtcacttctatcccttcagtcagtcagtcagtcagtcagtcagtcagtc 504
Db 105 TCCCTCTGGCAGTCACCTTCTATCCCTTCAGTGAGCAATGCTTTAACTGGAGAGAAATTC 46
Qy 505 agttttatcagt 517
Db 45 AGTTTATTCAGT 33

RESULT 4
AX140725/c
LOCUS AX140725
DEFINITION Sequence 215 from Patent WO0134802.
ACCESSION AX140725
VERSION AX140725.1 GI:14280836
KEYWORDS human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 366)
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,
Skeiky,Y.A. and Wang,A.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0134802-A 215 17-MAY-2001;
CORIXA CORPORATION (US)
```

```
FEATURES
source Location/Qualifiers
1. 366
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 119 a 79 c 69 g 96 t 3 others
ORIGIN
Query Match 36.1%; Score 202; DB 6; Length 366;
Best Local Similarity 99.6%; Pred. No. 3.4e-101;
Matches 252; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 265 ttccgtatggtccatgttgcttacagcctctgctaccgatgagaaggtcagagagatat 324
Db 285 TTTCGTATGGTCCATGTTGGCTTACAGCCTCTGCTTACCGATGAGAAGGTTCAGAGAGATAT 226
Qy 325 ttgtttctcaacatggtttatcagcaggttcatgcataattgaaactcttgggaatgag 384
Db 225 TTGTTTCTCAACATGGCTTATCAGCAGGTTTCATGCAATATTTGAAACTCTTGGGAATGAG 166
Qy 385 gaagaagtttgagaattgaaatgtatctctcttgccataatgagccttggcttactt 444
Db 165 GAAGAAGTTTGAGAAATGAAATGTATATNTCCCTTTGGCATAATGAGCCTTGGCTTACTT 106
Qy 445 tccctctggcagtcacttctatcccttcagtcagtcagtcagtcagtcagtcagtcagtc 504
Db 105 TCCCTCTGGCAGTCACCTTCTATCCCTTCAGTGAGCAATGCTTTAACTGGAGAGAAATTC 46
Qy 505 agttttatcagt 517
Db 45 AGTTTATTCAGT 33

RESULT 5
AX200585/c
LOCUS AX200585
DEFINITION Sequence 215 from Patent WO0151633.
ACCESSION AX200585
VERSION AX200585.1 GI:15390405
KEYWORDS human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 366)
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,
Stolk,J.A., Skeiky,Y.A., Wang,A. and Meagher,M.J.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0151633-A 215 19-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1. 366
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 119 a 79 c 69 g 96 t 3 others
ORIGIN
Query Match 36.1%; Score 202; DB 6; Length 366;
Best Local Similarity 99.6%; Pred. No. 3.4e-101;
Matches 252; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 265 ttccgtatggtccatgttgcttacagcctctgctaccgatgagaaggtcagagagatat 324
Db 285 TTTCGTATGGTCCATGTTGGCTTACAGCCTCTGCTTACCGATGAGAAGGTTCAGAGAGATAT 226
Qy 325 ttgtttctcaacatggtttatcagcaggttcatgcataattgaaactcttgggaatgag 384
Db 225 TTGTTTCTCAACATGGCTTATCAGCAGGTTTCATGCAATATTTGAAACTCTTGGGAATGAG 166
Qy 385 gaagaagtttgagaattgaaatgtatctctcttgccataatgagccttggcttactt 444
```

```
repeat_region complement(11315..11984)
repeat_region /rpt_family="L1"
repeat_region 11666..11687
repeat_region /rpt_family="L1"
repeat_region complement(12010..12299)
repeat_region /rpt_family="ALU"
repeat_region complement(12301..13893)
repeat_region /rpt_family="L1"
repeat_region 13897..14096
repeat_region /rpt_family="L1"
repeat_region complement(14777..14838)
repeat_region /rpt_family="L1"
misc_feature /note="similar to human EST T02878 (NID:g3141119)"
misc_feature 15715..15767
misc_feature /note="similar to human EST AA123941 (NID:g1682616)
misc_feature mq22f09.r1"
repeat_region complement(25113..25148)
repeat_region /rpt_family="L1"
repeat_region complement(25561..25578)
repeat_region /rpt_family="L1"
repeat_region complement(26832..27124)
repeat_region /rpt_family="ALU"
misc_feature 27125..27383
misc_feature /note="similar to human EST AA151807 (NID:g1720502)
misc_feature 27125..27544
misc_feature /note="similar to human EST AA151796 (NID:g1720491)
misc_feature 27125..27544
misc_feature /note="similar to human EST AA149579 (NID:g1720380)
misc_feature complement(27178..27600)
misc_feature /note="similar to human EST AA149579 (NID:g1720380)
misc_feature z139c02.s1"
misc_feature complement(27441..27890)
misc_feature /note="similar to human EST N52554 (NID:g1193720)
misc_feature yv36a11.s1"
misc_feature 27686..27856
misc_feature /note="similar to human EST W32120 (NID:g1313113)
misc_feature 27686..27838
misc_feature /note="similar to human EST W32085 (NID:g1313105)
misc_feature 2897b08.r1"
misc_feature complement(27733..28143)
misc_feature /note="similar to human EST N59831 (NID:g1203721)
misc_feature yz7a10.s1"
misc_feature complement(27838..28143)
misc_feature /note="similar to human EST W31561 (NID:g1312680)
misc_feature 2897b08.s1"
misc_feature complement(27887..28143)
misc_feature /note="similar to human EST N59830 (NID:g1203720)
misc_feature yz7a09.s1"
misc_feature complement(27997..28144)
misc_feature /note="similar to human EST W31628 (NID:g1312688)
misc_feature 28666..28972
misc_feature /note="similar to human EST R09339 (NID:g761262)
misc_feature yf26d05.r1"
misc_feature complement(29587..29853)
misc_feature /note="similar to human EST R09227 (NID:g761150)
misc_feature yf26d05.s1"
misc_feature 30146..30314
misc_feature /note="similar to human EST N78042 (NID:g1240743)
misc_feature yv71e06.r1"
misc_feature complement(31170..31350)
misc_feature /note="similar to human EST N58451 (NID:g1202341)
misc_feature yv71e06.s1"
repeat_region 31970..32006
repeat_region /rpt_family="L1"
repeat_region 32263..32563
repeat_region /rpt_family="ALU"
repeat_region complement(34737..34778)
repeat_region /rpt_family="L1"
repeat_region 36558..36709
repeat_region /rpt_family="ALU"
```

```
repeat_region complement(37630..37740)
repeat_region /rpt_family="PTR"
repeat_region 39839..40125
repeat_region /rpt_family="ALU"
repeat_region 40126..40157
repeat_region /rpt_family="L1"
repeat_region complement(41340..41706)
repeat_region /rpt_family="ALU"
repeat_region complement(41904..41933)
repeat_region /rpt_family="L1"
repeat_region 42348..42409
repeat_region /rpt_family="ALU"
repeat_region complement(42942..42982)
repeat_region /rpt_family="L1"
repeat_region complement(44380..44423)
repeat_region /rpt_family="L1"
repeat_region 44659..44954
repeat_region /rpt_family="ALU"
repeat_region 45422..45448
repeat_region /rpt_family="L1"
repeat_region complement(45615..45669)
repeat_region /rpt_family="L1"
repeat_region complement(45672..45963)
repeat_region /rpt_family="ALU"
repeat_region complement(45964..46597)
repeat_region /rpt_family="L1"
repeat_region 46100..46516
repeat_region /rpt_family="L1"
repeat_region 48370..48785
repeat_region /rpt_family="L1"
repeat_region complement(49577..50216)
repeat_region /rpt_family="L1"
repeat_region complement(50540..50791)
repeat_region /rpt_family="THE"
repeat_region complement(50820..50985)
repeat_region /rpt_family="THE"
repeat_region complement(51124..51156)
repeat_region /rpt_family="L1"
repeat_region 51405..51691
repeat_region /rpt_family="ALU"
repeat_region complement(51721..51764)
repeat_region /rpt_family="L1"
repeat_region complement(51765..52058)
repeat_region /rpt_family="ALU"
repeat_region complement(52252..52467)
```

Query Match 50.5%; Score 283; DB 9; Length 156214;
Best Local Similarity 99.7%; Pred. No. 1.7e-146;
Matches 333; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 20 ccttgcagagatgtgattccatcctatgctagaaacacacagagtgactttacaaaa 79
|||||
Db 19440 CCTTGTGCAGAGATGATTCATCCATATGCTAGAAACACAGAGTGACITTTACAAA 19499
|||||
QY 80 ttccatagagattgtaataaaaccttaccctatagttgccttacttctcctccctag 139
|||||
Db 19500 TTCCATATAGAGATGTAATAAAACCTTACCTATAGTTGCCATTACTTGTCTCCCTAG 19559
|||||
QY 140 tatacctcgcaggtctcttcggcagctgcttatcaactttattacggcacaagtatagga 199
|||||
Db 19560 TATACCTCGCAGGCTCTCTGGCAGCTGCTTATCAACTTTTATACGGCACCAGTATAGGA 19619
|||||
QY 200 gatttcacacctggtggaaacctggttacagtgtagaacaacagctggattactagtt 259
|||||
Db 19620 GATTTCCACCTTGGTGGAAACCTTGGTACAGTGTAGAAAACAGCTTGGATTACTAGTT 19679
|||||
QY 260 ttatcttcgctatggttcctatggttcctacagcctctgcttaccgcatgagaaggtcacaga 319
|||||
Db 19680 TTTTCTTCGCTATGTTCCATGTTGCCCTTACAGCCTCTCTTACCGATGAGAAGGTCACAGA 19739
|||||
QY 320 gatattgtttctcaacatgctgcttatcagcaggt 353
|||||
Db 19740 GATATTGTTTCTCAACATGGCTTATCAGCAGGT 19773
|||||
```

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC42678.1"

/db_xref="GI:14536714"

/translation="MESISMGPSPKLSSETCLPNCINGIKDKARKVTVCIGSGDPAKS

LTIRLCRGYHVYIGSRNPKFASEFFPHVDVTHREDALTKTNIFVAIHREHYTSLW
DLRHLLVKKILIDVSNMKRIQYIPESNAEYLASDFPSDLIVKGFNVSAWALQJGPKD
ASRYYICSNIRQAOVIELARQNFIPIDIGLSLSAREIEINPLRFTLWRGPPVV
AISLATFFFLSFVRDVHPYARNQSDYKIPIEIVANKHTLPPIVATITLLSVLYLAGLL
AAAYOLYGTGYRRPPWLEWLOCRKOLGILLSFFFAVHVHVASLCLPMRSELYLFL
NMAYQOVHANTENSWEEVEWRIEMYSIGFMSLGLLSLLAVTSIPSVSNALNWREFS
FIQSTLGIYVALLISTFHVLLIYGWKRFAEEYRYFPNPFVLLVLPISVILDLQLC
RIPD

BASE COUNT 648 a 537 c 520 g 748 t

Query Match 78.4%; Score 439; DB 6; Length 2453;

Best Local Similarity 99.6%; Pred. No. 2e-233;

Matches 539; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 20 ccttctgcagagatggtgattccatcatgctagaaacacacagagtgacttttcaaaa 79

Db 1043 CCTTTGTACAGATGATTCATCCATATGCTAGAACCAACAGAGTGACTTTTACAAA 1102

QY 80 ttctctagagattgtaataaaaccttacctatggtgaccttcttctctccctag 139

Db 1103 TTCCCTATAGAGATTGTGAATAAAACCTTACTATAGTTGCCATTACTTTGCTCTCCCTAG 1162

QY 140 tatactgcaggtcttctgcagctgcttcaactttattcagccaccagtagtaga 199

Db 1163 TATACCTTCAGGCTCTCTGGCAGCTGCTTATCACTTTATTACGGCACCAAGTATAGGA 1222

QY 200 gatttccaccttgggtgaaacctggttaccagtgtagaacaacagcttggtactactaagtt 259

Db 1223 GATTTCACCTTGGTGTGAACCTGGTTACAGTGTAGAAACAGCTTGGATTACTAAGTT 1282

QY 260 ttattctgctatggttccatggttgcctacagctctgtctaccgatgagaaggtcagaga 319

Db 1283 TTTTCTTCCTATGTGTCATGTGCTACAGGCTCTGCTTACCGATGAGAAGGTACAGGA 1342

QY 320 gatattgttctcaacatggtcttaccaggttccatgaggttccatgaggttccatgaggt 379

Db 1343 GATATTTGTTTCTACATGCTTATACAGAGTTCATGCAATATATGAAACTCTTGA 1402

QY 380 atgagaagaagtttggagaattgaaatgtatatctcttctgacataatgagccttggtc 439

Db 1403 ATGAGGAAGAAGTTTGGAGAATTGAAATGTATATCTCTTTGGCATAATGAGCCTTGGCT 1462

QY 440 tactttccctctgagtcacttctctccctcagtgagcaatgctttaaactggagag 499

Db 1463 TACTTTCCCTCTGGCAGTCACTTCTATCCCTTCAGTGAGCAATGCTTTAAACTGGAGAG 1522

QY 500 aattcagtttttaccagtcacacttgatagtgagctgctgctgctcataagtaactttccatg 559

Db 1523 AATTGAGTTTATTCAGTCTACACTTGGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1582

QY 560 t 560

Db 1583 T 1583

RESULT 2

LOCUS HSAC002064 156214 bp DNA linear PRI 09-MAY-1997

DEFINITION Human BAC clone RG016J04 from 7q21, complete sequence.

ACCESSION AC002064

VERSION AC002064.1 GI:2076723

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 156214)

TITLE Gattung,S. and Maggi,L.

JOURNAL The sequence of H. sapiens BAC clone RG016J04

AUTHORS Unpublished (1997)

TITLE 2 (bases 1 to 156214)

JOURNAL Waterston,R.

AUTHORS Direct Submission

TITLE Submitted (09-MAY-1997)

JOURNAL Genome Sequencing Center

COMMENT Department of Genetics, Washington University

St. Louis, MO 63108, USA

http://genome.wustl.edu/gsc

e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7> or send an E-mail to egreen@nhgri.nih.gov

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996). VECTOR: pBEO.

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of H_RG016J04; actual end is at 156214 of H_RG016J04. The orientation of this clone is unknown.

This clone contains STS SWSS2784 (NID:g1113580) and SWSS893 (NID:g454733).

FEATURES

source

Location/Qualifiers

1..156214

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21"

/clone="RG016J04"

/clone_lib="CITB-978SK-B"

15..40

/rpt_family="L1"

326..621

/rpt_family="ALU"

complement(977..1499)

/rpt_family="L1"

3398..3421

/rpt_family="L1"

5319..5345

/rpt_family="L1"

complement(8071..8347)

/rpt_family="ALU"

complement(9406..9975)

/rpt_family="L1"

complement(10000..11285)

/rpt_family="L1"

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:53:50 ; Search time 4356.39 Seconds
(without alignments)
2690.039 Million cell updates/sec

Title: US-09-802-520-6
Perfect score: 560
Sequence: 1 acattttttcttgatgc.....ctcataagtactttccatgt 560

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_em.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	439	78.4	2453	6	AX155249	AX155249	Sequence	Human BAC
2	283	50.5	156214	9	HSAC002064	AX106434	Sequence	Human BAC
3	202	36.1	366	6	AX106434	AX106434	Sequence	Human BAC
4	202	36.1	366	6	AX106434	AX106434	Sequence	Human BAC
5	202	36.1	366	6	AX106434	AX106434	Sequence	Human BAC
6	202	36.1	366	6	AX106434	AX106434	Sequence	Human BAC
7	191	34.1	162928	2	AC104475	AC104475	Sequence	Human BAC
8	174	31.1	322	6	AX155253	AX155253	Sequence	Human BAC
9	167	29.8	172915	2	AC099742	AC099742	Sequence	Human BAC
10	132	23.6	250	6	AX106432	AX106432	Sequence	Human BAC
11	132	23.6	250	6	AX106432	AX106432	Sequence	Human BAC
12	132	23.6	250	6	AX106432	AX106432	Sequence	Human BAC
13	132	23.6	250	6	AX106432	AX106432	Sequence	Human BAC
14	53	9.5	87401	2	AC021898	AC021898	Sequence	Human BAC
15	48	8.6	444	6	AX106433	AX106433	Sequence	Human BAC
16	48	8.6	444	6	AX106433	AX106433	Sequence	Human BAC
17	48	8.6	444	6	AX106433	AX106433	Sequence	Human BAC
18	48	8.6	444	6	AX106433	AX106433	Sequence	Human BAC
19	23	4.1	205085	2	AC026813	AC026813	Sequence	Human BAC
20	23	4.1	227144	2	AC092404	AC092404	Sequence	Human BAC
21	22	3.9	110000	9	HSY237C10_0	HSY237C10_0	Sequence	Human BAC
22	22	3.9	121084	2	AC010791	AC010791	Sequence	Human BAC
23	22	3.9	131002	2	AC024392	AC024392	Sequence	Human BAC
24	22	3.9	144645	9	AC026273	AC026273	Sequence	Human BAC
25	22	3.9	149577	9	AP000827	AP000827	Sequence	Human BAC
26	22	3.9	149745	9	AP000648	AP000648	Sequence	Human BAC
27	22	3.9	157666	9	AC034151	AC034151	Sequence	Human BAC
28	22	3.9	158187	9	AL512605	AL512605	Sequence	Human BAC
29	22	3.9	161973	9	HOAC002038	HOAC002038	Sequence	Human BAC
30	22	3.9	173005	2	AP001769	AP001769	Sequence	Human BAC
31	22	3.9	174876	2	AL359955	AL359955	Sequence	Human BAC
32	22	3.9	188872	9	AC020760	AC020760	Sequence	Human BAC
33	22	3.9	194487	2	AL591379	AL591379	Sequence	Human BAC
34	22	3.9	200149	9	AC006548	AC006548	Sequence	Human BAC
35	22	3.9	206192	9	AL133173	AL133173	Sequence	Human BAC
36	22	3.9	238351	2	AL445993	AL445993	Sequence	Human BAC
37	21	3.8	151976	2	AC015921	AC015921	Sequence	Human BAC
38	21	3.8	161049	2	AL591842	AL591842	Sequence	Human BAC
39	21	3.8	173767	9	AC002449	AC002449	Sequence	Human BAC
40	20	3.6	26539	3	U64836	U64836	Sequence	Human BAC
41	20	3.6	39582	3	U29244	U29244	Sequence	Human BAC
42	20	3.6	60932	2	AC105037	AC105037	Sequence	Human BAC
43	20	3.6	69663	2	AC090747	AC090747	Sequence	Human BAC
44	20	3.6	94487	8	AC012394	AC012394	Sequence	Human BAC
45	20	3.6	96000	9	AC079684	AC079684	Sequence	Human BAC

ALIGNMENTS

RESULT	1	AX155249	AX155249	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
LOCUS		AX155249	AX155249	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
DEFINITION		AX155249	AX155249	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
ACCESSION		AX155249	AX155249	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
VERSION		AX155249.1	AX155249.1	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
KEYWORDS		AX155249.1	AX155249.1	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
SOURCE		AX155249.1	AX155249.1	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
ORGANISM		AX155249.1	AX155249.1	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
REFERENCE		AX155249.1	AX155249.1	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
AUTHORS		AX155249.1	AX155249.1	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
TITLE		AX155249.1	AX155249.1	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
JOURNAL		AX155249.1	AX155249.1	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
FEATURES		AX155249.1	AX155249.1	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
source		AX155249.1	AX155249.1	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001
CDS		AX155249.1	AX155249.1	Sequence	7 from Patent WO0140276.	2453 bp	DNA	linear	PAT 22-JUN-2001

THIS PAGE BLANK (USPTO)

Search completed: September 20, 2002, 08:33:47
Job time: 8711 sec

```
RESULT 15
AZ045243
LOCUS
DEFINITION
  AZ045243 760 bp DNA linear GSS 08-MAR-2000
  nbeb0051M09f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
  clone OSJNB0051M09f, DNA sequence.
ACCESSION
  AZ045243
VERSION
  AZ045243.1 GI:7206979
KEYWORDS
  GSS.
SOURCE
  Oryza sativa.
  Oryza sativa
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Ehrhartoideae; Oryzaceae; Oryza.
  1 (bases 1 to 760)
  Wing, R.A. and Dean, R.A.
  A BAC End Sequencing Framework to Sequence the Rice Genome
  Unpublished (1998)
  Contact: Wing RA
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Tel: 864 656 7288
  Fax: 864 656 4293
  Email: rwing@clemson.edu
  Seq primer: GTAAACGACGCGCCAGTG
  Class: BAC ends
  High quality sequence start: 31
  High quality sequence stop: 304.
FEATURES
  Location/Qualifiers
    1..760
      /organism="Oryza sativa"
      /strain="Japonica"
      /cultivar="Nipponbare"
      /db_xref="taxon:4530"
      /clone="OSJNB0051M09f"
      /clone_lib="CUGI Rice BAC Library (EcoRI)"
      /tissue_type="Leaf"
      /lab_host="E. coli DH10B"
      /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
      Rice is the most important food crop in the world. Half of
      the world population, especially those inhabiting highly
      populated areas of the humid tropics and subtropics, rely
      on rice as their primary source of carbohydrate.
      Monocotyledonous rice is a diploid plant (2n=24) with a
      haploid genome equivalent of 431 Mbp (Arumuganathan and
      Earle, 1991). The relatively small genome of rice, three
      times larger than that of Arabidopsis, makes it suitable
      for genomic studies. In order to facilitate positional
      cloning, physical mapping and genome sequencing of rice,
      we have constructed a BAC library from Oryza sativa,
      Nipponbare variety using EcoRI as the cloning enzyme. The
      library contains 55,296 clones with an average insert size
      of 121 kb providing approximately 15 haploid genome
      equivalents. The deep coverage allows the isolation a
      particular sequence with a probability of 99.9 %. Three
      high density filters, each containing 18,432 clones
      (doubly spotted), represent the whole library for colony
      screening and can be requested from the Clemson University
      BAC/EST Resource Center (www.genome.clemson.edu)."
```

```
BASE COUNT      230 a      140 c      146 g      242 t
ORIGIN
```

```
Query Match      5.0%; Score 20; DB 12; Length 760;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy  82  ggtttatatgcagcaaca 101
    |||
Db   239  GGTTCATATATGCAGCAACA 258
```

. . .

Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu).

BASE COUNT 156 a 99 c 96 g 159 t 2 others
ORIGIN

Query Match 5.0%; Score 20; DB 12; Length 512;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 82 ggtttatatgcagcaaca 101
|||||
Db 242 GGTATATATGCAGCAACA 261

RESULT 13
A0831336
LOCUS
DEFINITION A0831336 566 bp DNA linear GSS 27-AUG-1999
genomic clone Plate-1097 Col-8 Row-A, DNA sequence.
ACCESSION A0831336
VERSION A0831336.1 GI:5797398
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 566)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc.Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380389
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (<http://www.resgen.com>). BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 1097 row: A column: 8
Seq primer: T7
Class: BAC ends
High quality sequence stop: 566.

FEATURES
source
1. .566
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate-1097 Col-8 Row-A"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"

/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"
BASE COUNT 183 a 84 c 73 g 215 t 11 others
ORIGIN

Query Match 5.0%; Score 20; DB 12; Length 566;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 247 tataagcttgccacatttt 266
|||||
Db 206 TATAGCTTGCCACATTTT 225

RESULT 14
AZ019656
LOCUS
DEFINITION AZ019656 721 bp DNA linear GSS 25-FEB-2000
RPCI-23-335E5-TV RPCI-23 Mus musculus genomic clone RPCI-23-335E5,
DNA sequence.
ACCESSION AZ019656
VERSION AZ019656.1 GI:7095040
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 721)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.

TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-335E5.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Research Genetics (<http://www.resgen.com>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 335 row: E column: 5
Seq primer: T7
Class: BAC ends.

FEATURES
source
1. .721
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-335E5"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 216 a 150 c 124 g 230 t 1 others
ORIGIN

Query Match 5.0%; Score 20; DB 12; Length 721;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 ctttgattgtcaaggattt 28
|||||
Db 94 CTTTGATTGTCAAAGGATT 113

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM10235 row: 1 column: 09
 High quality sequence stop: 639.

FEATURES

source

1. .838
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4450328"
 /clone_lib="NIH_MGC_90"
 /tissue_type="adenocarcinoma, cell_line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 245 a 154 c 172 g 267 t
 ORIGIN

Query Match 5.2%; Score 21; DB 10; Length 838;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 ttttttttcctttattccttt 283
 |||||
 Db 832 TTTTTCCTTATTCCTTT 812

RESULT 11

T87479
 LOCUS T87479 488 bp mRNA linear EST 17-MAR-1995
 DEFINITION Yd85e07 r1 Soares fetal liver spleen INFLS Homo sapiens CDNA clone
 IMAGE:115044 5', mRNA sequence.

ACCESSION T87479
 VERSION T87479.1 GI:715831
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 488)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
 M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
 R., Williamson, A., Wohldmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu

Insert Size: 1499

High quality sequence stops: 253 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1499 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 253.

Location/Qualifiers

FEATURES

source

1. .488
 /organism="Homo sapiens"
 /db_xref="GDB:470861"
 /db_xref="taxon:9606"
 /clone="IMAGE:115044"
 /clone_lib="Soares fetal liver spleen INFLS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"

/lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 15' AACCGAAGATTAAATAAGATCTTTTCTTTTCTTTT 3',
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Patima Bonaldo."

BASE COUNT 159 a 107 c 99 g 119 t
 ORIGIN

Query Match 5.0%; Score 20; DB 10; Length 488;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 ggccacatttttttcccttt 275
 |||||

Db 448 GGCCACATTTTTCCTTT 467

RESULT 12

AQ912082
 LOCUS nbeb0016Allf CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
 DEFINITION clone nbeb0016Allf, DNA sequence.

ACCESSION AQ912082

VERSION AQ912082.1 GI:6508598

KEYWORDS GSS.

SOURCE Oryza sativa.

ORGANISM

Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 512)

Wing, R.A. and Dean, R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: wing@clemson.edu

Seq primer: TAATACGACTCACTATAGGG

Class: BAC ends

High quality sequence start: 30

High quality sequence stop: 407.

FEATURES

source

1. .512
 /organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone="nbeb0016Allf"
 /clone_lib="CUGI Rice BAC Library (EcoRI)"
 /tissue_type="Leaf"
 /lab_host="E. coli DH10B"

/note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
 Rice is the most important food crop in the world. Half of
 the world population, especially those inhabiting highly
 populated areas of the humid tropics and subtropics, rely
 on rice as their primary source of carbohydrate.

Monocotyledonous rice is a diploid plant (2n=24) with a
 haploid genome equivalent of 431 Mbp (Arumuganathan and
 Earle, 1991). The relatively small genome of rice, three
 times larger than that of Arabidopsis, makes it suitable
 for genomic studies. In order to facilitate positional
 cloning, physical mapping and genome sequencing of rice,
 we have constructed a BAC library from Oryza sativa,

FEATURES
source

Location/Qualifiers
1. 561
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-brx-g-01-0-UI"
/clone.lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged, normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCR amplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_L1B=NIH_BMAP_M_S4
TAG_TISSUE=pineal-glands
TAG_SEQ=CAGAC"

BASE COUNT 154 a 115 c 129 g 163 t
ORIGIN

Query Match 6.5%; Score 26; DB 10; Length 561;

Best Local Similarity 100.0%; Pred. No. 0.054;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 caggtttatatgtcagcaacaatat 105

Db 250 CAGGTTTATATATGCAGCAACAATAT 225

RESULT 7

AZ428664/c
LOCUS 712 bp DNA linear GSS 03-OCT-2000
DEFINITION IM0212F14F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0212F14 F, DNA sequence.
ACCESSION AZ428664
VERSION AZ428664.1 GI:10552677
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 712)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center

University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0212 row: F column: 14
Seq primer: CTTGTAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 712.
Location/Qualifiers
1. 712

FEATURES
source

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0212F14"
/clone.lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 223 a 160 c 146 g 183 t
ORIGIN

Query Match 5.7%; Score 23; DB 12; Length 712;

Best Local Similarity 100.0%; Pred. No. 1.5;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ttgattgtcaaggatttaagt 33

Db 98 TTGATTGTCAAAGGATTAATGT 76

RESULT 8

LOCUS BB621516/c
DEFINITION BB621516 RIKEN full-length enriched, 13 days embryo male testis Mus musculus cDNA clone 6030497E09 5', mRNA sequence.
ACCESSION BB621516
VERSION BB621516.1 GI:16460206
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 628)

AUTHORS Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, T., Muramatsu, M. and Hayashizaki, Y.

TITLE RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

AA508880
 AA508880.1 GI:2245821
 EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 1 (bases 1 to 322)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 M.D., Michael R. Emmert-Buck, M.D., Ph.D., Rodrigo F. Chuvaqui,
 cDNA Library Preparation: David B. Krizman, Ph.D.
 DNA sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bhrp/image/image.html
 Insert length: 1000 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 249.

FEATURES
 source
 1..322
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:941693"
 /clone_lib="NCI_CGAP_Pr6"
 /sex="male"
 /tissue_type="prostate"
 /lab_host="DH10B"
 /note="vector: PAMp10; mRNA made from prostatic
 intraepithelial neoplasia (low-grade), cDNA made by
 oligo-dT priming. Non-directionally cloned.
 Size-selected on agarose gel, average insert size 600 bp.
 Reference: Krizman et al. (1996) Cancer Research
 56:5380-5383."

BASE COUNT 84 a 70 c 62 g 106 t

ORIGIN

Query Match 12.4%; Score 50; DB 9; Length 322;
 Best Local Similarity 100.0%; Pred. No. 1.3e-13;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 342 ctatagagattgtgataaaaccttacctatagtgtgcattcttc 391
 |||||
 Db 83 CTATAGAGATTGTGAATAAACCTTACCTATAGTTGCACTTCTGCTC 132

RESULT 5
 LOCUS BI359597 595 bp mRNA linear EST 01-AUG-2001
 DEFINITION 384188 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
 ACCESSION BI359597
 VERSION BI359597.1 GI:15055625
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa

REFERENCE
 1 (bases 1 to 595)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
 Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
 and Keele,J.W.
 Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine
 Unpublished (2000)
 CONTACT: Smith TPL
 USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGCAACAGCTATCACCAT
 BACKWARD: GTTTCCTCCAGTCACGACG
 Plate: 132 row: H column: 3
 Seq primer: ATTAGGTGACACTATAG.
 Location/Qualifiers
 1..595
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 2PIG"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from testis, ovary,
 endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 143 a 138 c 126 g 188 t

ORIGIN

Query Match 7.9%; Score 32; DB 10; Length 595;
 Best Local Similarity 100.0%; Pred. No. 6.5e-05;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 215 ttactctctgagagggccagtggtgtagc 246
 |||||
 Db 88 TTTACTCTGGAGAGGGCCAGTGGTGTAGC 119

RESULT 6
 BI133492/c
 LOCUS BI133492 561 bp mRNA linear EST 02-JUL-2001
 DEFINITION UI-M-BH3-brx-g-01-0-UI.sl NIH_BMAP_M_S4 Mus musculus cDNA clone
 UI-M-BH3-brx-g-01-0-UI 3', mRNA sequence.
 ACCESSION BI133492
 VERSION BI133492.1 GI:14583740
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 1 (bases 1 to 561)
 Bernaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 CONTACT: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mest@mail.nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 pineal glands tissue cDNA Library Preparation: M.B. Soares Lab
 Clone distribution: Researchers may obtain BMAP cDNA clones from
 RESEARCH GENETICS. It should be noted that Bento Soares is
 generating a small number of additional specialized non-redundant
 arrays of BMAP cDNAs whose availability will be considered under
 appropriate and limited collaborative arrangements
 Seq primer: M13 Forward
 POLYA=Yes.

```

/db_xref="taxon:9606"
/dev_stage="Adult"
/site="Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORSTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      144 a  100 c   99 g   140 t
ORIGIN

Query Match      32.0%; Score 129; DB 10; Length 483;
Best Local Similarity 100.0%; Pred. No. 5e-52;
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cccagattcttggatgcacaaaggatttaattgtctcagcttggcgaccttcagttagg 60
|||||
Db 220 CCCAGATTCTTTGATGTCAAGGATTAAATGTTCTCAGCTTGGGCACTTCAGTTAGG 161
|||||

Qy 61 acctaaagatccagccgagctttatatgcagcaacaattatcagcgcgacaaca 120
|||||
Db 160 ACCTAGATGCCAGCCGCGAGTTTATATATGACACAAATATTCAGCGCGACAACA 101
|||||

Qy 121 ggtttatga 129
|||||
Db 100 GGTATTGA 92

RESULT 2
AA888667
LOCUS      nw77b11.s1 NCI_CGAP_Pr12 Homo sapiens cDNA clone IMAGE:1252605,
DEFINITION mRNA sequence.
ACCESSION AA888667
VERSION   AA888667.1 GI:3004342
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 180)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,
          Rodrigo F. Chuquai, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
          cDNA Library Preparation: David B. Krizman, Ph.D.
          cDNA Library Arrayed by: Greg Lennon, Ph.D.
          DNA Sequencing by: Washington University Genome Sequencing Center
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          www-bio.llnl.gov/bbrp/image/image.html
          Seq primer: -40ml3 fwd. ET from Amersham.
FEATURES
Source      Location/Qualifiers
1..180
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1252605"
/clone_lib="NCI_CGAP_Pr12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="vector: pAMP10; mRNA made from metastatic prostate
lesion of the bone, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Krizman,
NIH."

/db_xref="taxon:9606"
/dev_stage="Adult"
/site="Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORSTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      144 a  100 c   99 g   140 t
ORIGIN

Query Match      32.0%; Score 129; DB 10; Length 483;
Best Local Similarity 100.0%; Pred. No. 5e-52;
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cccagattcttggatgcacaaaggatttaattgtctcagcttggcgaccttcagttagg 60
|||||
Db 220 CCCAGATTCTTTGATGTCAAGGATTAAATGTTCTCAGCTTGGGCACTTCAGTTAGG 161
|||||

Qy 61 acctaaagatccagccgagctttatatgcagcaacaattatcagcgcgacaaca 120
|||||
Db 160 ACCTAGATGCCAGCCGCGAGTTTATATATGACACAAATATTCAGCGCGACAACA 101
|||||

Qy 121 ggtttatga 129
|||||
Db 100 GGTATTGA 92

RESULT 2
AA888667
LOCUS      nw77b11.s1 NCI_CGAP_Pr12 Homo sapiens cDNA clone IMAGE:1252605,
DEFINITION mRNA sequence.
ACCESSION AA888667
VERSION   AA888667.1 GI:3004342
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 180)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,
          Rodrigo F. Chuquai, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
          cDNA Library Preparation: David B. Krizman, Ph.D.
          cDNA Library Arrayed by: Greg Lennon, Ph.D.
          DNA Sequencing by: Washington University Genome Sequencing Center
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          www-bio.llnl.gov/bbrp/image/image.html
          Seq primer: -40ml3 fwd. ET from Amersham.
FEATURES
Source      Location/Qualifiers
1..180
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1252605"
/clone_lib="NCI_CGAP_Pr12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="vector: pAMP10; mRNA made from metastatic prostate
lesion of the bone, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Krizman,
NIH."

```

```

BASE COUNT      56 a   31 c   37 g   56 t
ORIGIN

Query Match      12.9%; Score 52; DB 9; Length 180;
Best Local Similarity 100.0%; Pred. No. 1.5e-14;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 80 caggtttatatgcagcaacaattatcagcgcgacaacaggtttatgaac 131
|||||
Db 129 CAGGTTTATATGCGCAACAATATTTCAGCGCGACAACAGGTTATTGAAC 180
|||||

RESULT 3
AA935972
LOCUS      nv30h06.s1 NCI_CGAP_Pr12 Homo sapiens cDNA clone IMAGE:1273307,
DEFINITION mRNA sequence.
ACCESSION AA935972
VERSION   AA935972.1 GI:3093129
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 181)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,
          Rodrigo F. Chuquai, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
          cDNA Library Preparation: David B. Krizman, Ph.D.
          cDNA Library Arrayed by: Greg Lennon, Ph.D.
          DNA Sequencing by: Washington University Genome Sequencing Center
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          www-bio.llnl.gov/bbrp/image/image.html
          Seq primer: -40ml3 fwd. ET from Amersham.
FEATURES
Source      Location/Qualifiers
1..181
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1273307"
/clone_lib="NCI_CGAP_Pr12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="vector: pAMP10; mRNA made from metastatic prostate
lesion of the bone, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Krizman,
NIH."

BASE COUNT      56 a   34 c   35 g   56 t
ORIGIN

Query Match      12.9%; Score 52; DB 9; Length 181;
Best Local Similarity 100.0%; Pred. No. 1.5e-14;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 80 caggtttatatgcagcaacaattatcagcgcgacaacaggtttatgaac 131
|||||
Db 130 CAGGTTTATATGCGCAACAATATTTCAGCGCGACAACAGGTTATTGAAC 181
|||||

RESULT 4
AA508880
LOCUS      ng6f03.s1 NCI_CGAP_Pr6 Homo sapiens cDNA clone IMAGE:941693, mRNA
DEFINITION sequence.

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 08:33:44 ; Search time 5053.42 Seconds
(without alignments)
1076.355 Million cell updates/sec

Title: US-09-802-520-5
Perfect score: 403
Sequence: 1 ccagattctttagtgca.....cttgctcccttagtatac 403

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_estl:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	129	32.0	483	10 BF883142	BF883142 QV3-ET021
2	52	12.9	180	9 AA888667	AA888667 nw7b11.s
3	52	12.9	181	9 AA935972	AA935972 ny30h06.s
4	50	12.4	322	9 AA508880	AA508880 ng86f03.s
5	32	7.9	595	10 BI359597	BI359597 384188 MA
C 6	26	6.5	561	10 BI133492	BI133492 UI-N-BH3-
7	23	5.7	712	12 A2428664	A2428664 LM0212F14
C 8	22	5.5	628	9 BB621516	BB621516 BB621516
9	21	5.2	756	9 BE039135	BE039135 AB14A10 A
C 10	21	5.2	838	10 BG119803	BG119803 602352138
11	20	5.0	488	10 T87479	T87479 yd85e07.r1
12	20	5.0	512	12 AQ912082	AQ912082 nbe0016A
13	20	5.0	566	12 AQ831336	AQ831336 HS_5521_A
14	20	5.0	721	12 AZ019656	AZ019656 RPCI-23-3
15	20	5.0	760	12 A2045243	A2045243 nbe0051M
16	20	5.0	803	12 AG165588	AG165588 Pan trogl
17	19	4.7	109	12 AQ248591	AQ248591 T5K8-T7 T

C 18	19	4.7	210	9 AI305509	AI305509 qw72a10.x
C 19	19	4.7	214	9 AI223600	AI223600 qx31b05.x
C 20	19	4.7	250	9 AV264258	AV264258 AV264258
21	19	4.7	343	9 AU077687	AU077687 AU077687
C 22	19	4.7	391	10 BF895019	BF895019 QV1-WF016
C 23	19	4.7	396	10 BG019784	BG019784 GC68A09.X
24	19	4.7	431	10 T95226	T95226 ye44f06.r1
C 25	19	4.7	496	9 AA701826	AA701826 ET2271 Tr
C 26	19	4.7	505	10 BI350635	BI350635 fr34h06.Y
C 27	19	4.7	518	12 AQ647805	AQ647805 RPCI93-DP
C 28	19	4.7	544	12 AQ920208	AQ920208 RPCI-23-2
C 29	19	4.7	548	12 BH454179	BH454179 BOHON40TR
C 30	19	4.7	560	10 BM034788	BM034788 fu35q08.Y
C 31	19	4.7	563	10 BI845092	BI845092 fg35e10.Y
C 32	19	4.7	563	10 BM035184	BM035184 fu40e11.Y
C 33	19	4.7	565	12 AZ179970	AZ179970 SP_0167.A
C 34	19	4.7	577	10 BI864123	BI864123 ft15804.Y
C 35	19	4.7	579	10 BI864118	BI864118 ft15807.Y
C 36	19	4.7	614	10 BI472993	BI472993 fr94h01.Y
C 37	19	4.7	636	10 BI839566	BI839566 fg39h04.Y
C 38	19	4.7	644	10 BI428484	BI428484 fr84e12.Y
C 39	19	4.7	652	9 AW186789	AW186789 BNLGH1510
C 40	19	4.7	655	9 AV264339	AV264339 AV264339
C 41	19	4.7	659	9 AW305417	AW305417 fj59b06.Y
C 42	19	4.7	702	12 AZ371568	AZ371568 IM0123K05
C 43	19	4.7	704	9 AJ395679	AJ395679 AJ395679
C 44	19	4.7	720	10 BI704176	BI704176 zfshtARAM
C 45	19	4.7	779	9 AL584288	AL584288 AL584288

ALIGNMENTS

RESULT 1
BF883142/c
LOCUS QV3-ET0211-071200-529-b10 ET0211 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF883142
ACCESSION BF883142
VERSION BF883142.1 GI:12273268
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 483)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl-QV3&t2-QV3-ET0211-071200-529-b10&t3-2000-12-07&t4-1)
Seq primer: puc 18 forward
High quality sequence start: 26
High quality sequence stop: 483.
Location/Qualifiers
1. .483
/organism="Homo sapiens"

THIS PAGE BLANK (USPTO)

US-08-789-329C-6
 ; Sequence 6, Application US/08789329C
 ; Patent No. 6165755
 ; GENERAL INFORMATION:
 ; APPLICANT: SHERWOOD ET AL.
 ; TITLE OF INVENTION: CHICKEN NEUROPEPTIDE GENE USEFUL
 ; TITLE OF INVENTION: FOR IMPROVED POULTRY PRODUCTION
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Klarquist Sparkman Campbell Leigh &
 ; ADDRESSEE: Winston, LLP
 ; STREET: One World Trade Center
 ; STREET: 121 S.W. Salmon Street
 ; STREET: Suite 1600
 ; CITY: Portland
 ; STATE: Oregon
 ; COUNTRY: United States of America
 ; ZIP: 97204-2988
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Disk, 3-1/2 inch
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: Windows NT
 ; SOFTWARE: Wordperfect 7.0 & ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/789,329C
 ; FILING DATE: 01/23/97
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Earp, David J.
 ; REGISTRATION NUMBER: 41,401
 ; REFERENCE/DOCKET NUMBER: 2847-46468/DJE
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (503) 226-7391
 ; TELEFAX: (503) 228-9446
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1079 bp
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; US-08-789-329C-6

Query Match 4.08; Score 16; DB 4; Length 1079;
 Best Local Similarity 100.0%; Pred. NO. 37;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 263 ttttttttcctttatt 278
 ||||||||||||
 Db 21 TTTTTCCTTTATT 36

Search completed: September 20, 2002, 09:51:09
 Job time: 11857 sec

; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,883
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-883-56

Query Match 4.0%; Score 16; DB 1; Length 837;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 taaaaccttacctata 373
|||||
Db 627 TAAACCTTACCTATA 612

RESULT 13

US-08-832-877-56/c
; Sequence 56, Application US/08832877
; Patent No. 5840506
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 116
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, CONDA, LAVORGNA & MONACO, P.C.
; STREET: Suite 1800 Two Penn Center Plaza
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,877
; FILING DATE:
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-877-56

Query Match 4.0%; Score 16; DB 2; Length 837;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 taaaaccttacctata 373
|||||
Db 627 TAAACCTTACCTATA 612

RESULT 14

US-08-789-329C-9
; Sequence 9, Application US/08789329C
; Patent No. 6165755
; GENERAL INFORMATION:
; APPLICANT: SHERWOOD ET AL.
; TITLE OF INVENTION: CHICKEN NEUROPEPTIDE GENE USEFUL
; TITLE OF INVENTION: FOR IMPROVED POULTRY PRODUCTION
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klarquist Sparkman Campbell Leigh &
; ADDRESSEE: Whinston, LLP
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204-2988
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows NT
; SOFTWARE: WordPerfect 7.0 & ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/789,329C
; FILING DATE: 01/23/97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Earp, David J.
; REGISTRATION NUMBER: 41,401
; REFERENCE/DOCKET NUMBER: 2847-46468/DJE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 992 bp
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-789-329C-9

Query Match 4.0%; Score 16; DB 4; Length 992;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 ttttttttctttatt 278
|||||
Db 21 TTTTTCCTTATT 36

RESULT 15

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 178 atcagccagagagatt 193
 |||||
 Db 121 ATCAGCCAGAGAGATT 136

RESULT 10
 US-08-313-075A-35
 ; Sequence 35, Application US/08313075A
 ; Patent No. 5639870
 ; GENERAL INFORMATION:
 ; APPLICANT: Holton, Timothy A.
 ; APPLICANT: Cornish, Edwina C.
 ; APPLICANT: Tanaka, Yoshikazu
 ; APPLICANT: Lester, Diane R.
 ; TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOID
 ; TITLE OF INVENTION: PATHWAY ENZYMES AND USES THEREFOR
 ; NUMBER OF SEQUENCES: 58
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Scully, Scott, Murphy & Presser
 ; STREET: 400 Garden City Plaza
 ; CITY: Garden City
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 11530
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/313,075A
 ; FILING DATE: 30-NOV-1994
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: AU PL 1538/92
 ; FILING DATE: 27-MAR-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: AU PL 6698/93
 ; FILING DATE: 07-JAN-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: AU PCT/AU93/00127
 ; FILING DATE: 25-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Digiglio, Frank S.
 ; REGISTRATION NUMBER: 31,346
 ; REFERENCE/DOCKET NUMBER: 9433
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (516) 742-4343
 ; TELEFAX: (516) 742-4366
 ; TELEX: 230 901 SANS UR
 ; INFORMATION FOR SEQ ID NO: 35:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 203 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 3..203
 US-08-313-075A-35

Query Match 4.0%; Score 16; DB 1; Length 203;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 178 atcagccagagagatt 193
 |||||
 Db 121 ATCAGCCAGAGAGATT 136

RESULT 11
 US-08-502-046-27
 ; Sequence 27, Application US/08502046
 ; Patent No. 5861487
 ; GENERAL INFORMATION:
 ; APPLICANT: Holton, Timothy A.
 ; APPLICANT: Cornish, Edwina C.
 ; APPLICANT: Kovacic, Filippa
 ; APPLICANT: Tanaka, Yoshikazu
 ; APPLICANT: Lester, Diane R.
 ; TITLE OF INVENTION: GENETIC SEQUENCES ENCODING A 3,5'-
 ; TITLE OF INVENTION: HYDROXYLASE AND USES
 ; NUMBER OF SEQUENCES: 29
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Scully, Scott, Murphy & Presser
 ; STREET: 400 Garden City Plaza
 ; CITY: Garden City
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 11530
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/502,046
 ; FILING DATE: 14-JUL-1995
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/285,309
 ; FILING DATE: 03-AUG-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Digiglio, Frank S.
 ; REGISTRATION NUMBER: 31,346
 ; REFERENCE/DOCKET NUMBER: 86332
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (516) 742-4343
 ; TELEFAX: (516) 742-4366
 ; TELEX: 230 901 SANS UR
 ; INFORMATION FOR SEQ ID NO: 27:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 203 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 US-08-502-046-27

Query Match 4.0%; Score 16; DB 2; Length 203;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 178 atcagccagagagatt 193
 |||||
 Db 121 ATCAGCCAGAGAGATT 136

RESULT 12
 US-08-832-883-56/c
 ; Sequence 56, Application US/08832883
 ; Patent No. 5807681
 ; GENERAL INFORMATION:
 ; APPLICANT: Giordano, Antonio
 ; APPLICANT: Baidi, Alphonso
 ; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS
 ; TITLE OF INVENTION: OF CANCER
 ; NUMBER OF SEQUENCES: 115
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEIDEL, GONDA, LAVORNA & MONACO, P.C.
 ; STREET: Suite 1800 Two Penn Center Plaza
 ; CITY: Philadelphia

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313,075A
FILING DATE: 30-NOV-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PL 1538/92
FILING DATE: 27-MAR-1992
APPLICATION DATA:
APPLICATION NUMBER: AU PL 6698/93
FILING DATE: 07-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU93/00127
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 9433
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 180 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-313-075A-40

Query Match 4.0%; Score 16; DB 1; Length 180;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 atcagccagagagatt 193
|||||
DB 27 ATCAGCCAGAGATT 42

RESULT 8
US-07-912-900-27
Sequence 27, Application US/07912900
Patent No. 5349125
GENERAL INFORMATION:
APPLICANT: Holton, Timothy A.
APPLICANT: Cornish, Edwina C.
APPLICANT: Kovacic, Filippa
APPLICANT: Tanaka, Yoshikazu
APPLICANT: Lester, Diane R.
TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOID
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/912,900
FILING DATE: 19920713
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346

REFERENCE/DOCKET NUMBER: 8633
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-912-900-27

Query Match 4.0%; Score 16; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 atcagccagagagatt 193
|||||
DB 121 ATCAGCCAGAGATT 136

RESULT 9
US-08-285-309-27
Sequence 27, Application US/08285309
Patent No. 5569832
GENERAL INFORMATION:
APPLICANT: Holton, Timothy A.
APPLICANT: Cornish, Edwina C.
APPLICANT: Kovacic, Filippa
APPLICANT: Tanaka, Yoshikazu
APPLICANT: Lester, Diane R.
TITLE OF INVENTION: GENETIC SEQUENCES ENCODING A 3,5'-
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/285,309
FILING DATE: 03-AUG-1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 86332
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-285-309-27

Query Match 4.0%; Score 16; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 38;

```

RESULT      5
US-09-323-873A-9
  Sequence 9, Application US/09323873A
  Patent No. 6329503
  GENERAL INFORMATION:
  APPLICANT: Daniel E. Afar
  APPLICANT: Rene S. Hubert
  APPLICANT: Kahlan Leong
  APPLICANT: Arthur B. Raitano
  APPLICANT: Douglas C. Saffran
  APPLICANT: Steve Chappell Mitchell
  TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
  TITLE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF
  FILE REFERENCE: 129,16USU2
  CURRENT APPLICATION NUMBER: US/09/323,873A
  CURRENT FILING DATE: 1999-06-01
  PRIOR APPLICATION NUMBER: 60/087,520
  PRIOR FILING DATE: 1998-06-01
  PRIOR APPLICATION NUMBER: 60/091,183
  PRIOR FILING DATE: 1998-06-30
  NUMBER OF SEQ ID NOS: 32
  SOFTWARE: FastSeq for Windows Version 4.0
  SEQ ID NO 9
  LENGTH: 322
  TYPE: DNA

```

```

; ORGANISM: Homo Sapiens
US-09-323-873A-9

Query Match      12.4%  Score 50; DB 4; Length 322;
Best Local Similarity 100.0%; Pred. No. 9.7e-17;
Matches 50; Conservative 0; Mismatches 0; Indels 0;

QY 342 ctatagagattgtgaataaaacccctaacctatagtgtgccattacttgtctc 391
|||||
Db 83 ctatagagattgtgaataaaacccctaacctatagtgtgccattacttgtctc 132
|||||

RESULT 6
US-09-323-873A-7
; Sequence 7, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; FILE REFERENCE: 129.16US02
; CURRENT APPLICATION NUMBER: US/09/323.873A
; CURRENT FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-323-873A-7

Query Match      12.4%  Score 50; DB 4; Length 521;
Best Local Similarity 100.0%; Pred. No. 9.6e-17;
Matches 50; Conservative 0; Mismatches 0; Indels 0;

QY 342 ctatagagattgtgaataaaacccctaacctatagtgtgccattacttgtctc 391
|||||
Db 18 ctatagagattgtgaataaaacccctaacctatagtgtgccattacttgtctc 67
|||||

RESULT 7
US-08-313-075A-40
; Sequence 40, Application US/08313075A
; Patent No. 5639870
; GENERAL INFORMATION:
; APPLICANT: Holton, Timothy A.
; APPLICANT: Cornish, Edwina C.
; APPLICANT: Tanaka, Yoshikazu
; TITLE OF INVENTION: GENETIC SEQUENCES ENCODING FLAVONOID
; TITLE OF INVENTION: PATHWAY ENZYMES AND USES THEREFOR
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```


THIS PAGE BLANK (USPTO)

CC polynucleotide of the invention.

XX Sequence 444 BP; 120 A; 100 C; 78 G; 143 T; 3 other;

Query Match 54.3%; Score 219; DB 22; Length 444;

Best Local Similarity 99.6%; Pred. No. 2.le-104;

Matches 269; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccagattcttgattgtcaaaaggatttaattgttctcagcttgggcacttcagttagg 60

|||||

Db 38 ccagattcttgattgtcaaaaggatttaattgttctcagcttgggcacttcagttagg 97

|||||

QY 61 acctaaggatgccagccgaggtttatatatgcagcaacaataattcaagcgcgacaaca 120

|||||

Db 98 acctaaggatgccagccgaggtttatatatgcagcaacaataattcaagcgcgacaaca 157

|||||

QY 121 gttattgaacttgcgcgcaggttgaatttcattccattgacttgggatccttatcatc 180

|||||

Db 158 ggtattgaacttgcgcgcaggttgaatttcattccattgacttgggatccttatcatc 217

|||||

QY 181 agccagagagattgaaattaccctacgactcttactctctgagagggccagtggt 240

|||||

Db 218 agccanagagattgaaattaccctacgactcttactctctgagagggccagtggt 277

QY 241 ggtagctataagcttggccacattttttt 270

|||||

Db 278 ggtagctataagcttggccacattttttt 307

RESULT 15

AAS10206

ID AAS10206 standard; cDNA; 444 BP.

XX AC AAS10206;

XX AC AAS10206;

XX AC AAS10206;

DT 24-OCT-2001 (first entry)

XX Human prostate tumour cDNA 8-b6 #1.

DE Human; prostate tumour protein; prostate cancer; ss.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

QY 203 ccctacgactttactctctgagagggccagtggtgtagctataagcttggccaca 262
 |||||
 Db 121 ccctacgactttactctctgagagggccagtggtgtagctataagcttggccaca 180
 |||||
 QY 263 ttttttttcccttattcccttctgacagagtggtgattcccatatgctagaaa 315
 |||||
 Db 181 ttttttttcccttattcccttctgacagagtggtgattcccatatgctagaaa 233
 |||||

RESULT 11
 AAV61348
 ID AAV61348 standard; cDNA; 444 BP.
 XX
 AC AAV61348;
 XX
 DT 06-JAN-1999 (first entry)
 XX
 DE Forward DNA sequence of prostate tumour clone 8-b6.
 XX
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9837093-A2.
 XX
 PD 27-AUG-1998.
 XX
 PF 25-FEB-1998; 98WO-US03492.
 XX
 PR 09-FEB-1998; 98US-0020956.
 XX
 PR 25-FEB-1997; 97US-0806099.
 XX
 PR 01-AUG-1997; 97US-0904804.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Dillon DC, Xu J;
 XX
 DR WPI; 1998-609886/51.
 XX
 PT Polypeptides comprising immunogenic portions of prostate proteins -
 used in a vaccine for the treatment of prostate cancer
 XX
 PS Claim 12; Page 121; 130pp; English.
 XX
 CC The present sequence is a DNA which encodes an immunogenic portion
 of a prostate tumour protein. The encoded immunogen, or the DNA itself,
 CC can be used as a vaccine for the treatment of prostate cancer. The DNA
 CC was identified by analysis of a subtracted cDNA library obtained by
 CC subtracting a prostate tumour cDNA expression library with a normal
 CC tissue cDNA library.
 XX
 SQ Sequence 444 BP; 120 A; 100 C; 78 G; 143 T; 3 other;

Query Match 54.3%; Score 219; DB 19; Length 444;
 Best Local Similarity 99.6%; Pred. No. 2.le-104;
 Matches 269; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccagattcttgattgtcaaaaggatttaattgttctcagcttgggcacttcagttagg 60
 |||||
 Db 38 ccagattcttgattgtcaaaaggatttaattgttctcagcttgggcacttcagttagg 97
 |||||
 QY 61 acctaaggatgcagccgaggtttatatatgcagcaacaattattcaagcgcgacaaca 120
 |||||
 Db 98 acctaaggatgcagccgaggtttatatatgcagcaacaattattcaagcgcgacaaca 157
 |||||
 QY 121 ggatttgaacttgcgcgcagttgaatttcattccattgacttgggattcccttatc 180
 |||||
 Db 158 ggatttgaacttgcgcgcagttgaatttcattccattgacttgggattcccttatc 217
 |||||
 QY 181 agccagagagattgaaaaattaccctacgactcttactctctgtgagagggccagtggt 240
 |||||

Db 218 agccanagagattgaaaaattaccctacgactcttactctctgtgagagggccagtggt 277
 |||||
 QY 241 gtagctataagcttggccacattttttt 270
 |||||
 Db 278 gtagctataagcttggccacattttttt 307
 |||||

RESULT 12
 AAV58684
 ID AAV58684 standard; cDNA; 444 BP.
 XX
 AC AAV58684;
 XX
 DT 08-DEC-1998 (first entry)
 XX
 DE Prostate tumour specific gene clone 8-b6.
 XX
 KW Prostate tumour specific gene; human; prostate cancer; detection;
 KW therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9837418-A2.
 XX
 PD 27-AUG-1998.
 XX
 PF 25-FEB-1998; 98WO-US03690.
 XX
 PR 09-FEB-1998; 98US-0904809.
 XX
 PR 25-FEB-1997; 97US-0806596.
 XX
 PR 01-AUG-1997; 97US-0904809.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Dillon DC, Xu J;
 XX
 DR WPI; 1998-480805/41.
 XX
 PT Novel human prostate specific tumour protein and fragments - useful
 for detecting and treating prostate cancers
 XX
 PS Claim 1; Page 131; 141pp; English.
 XX
 CC This sequence represents a human prostate tumour specific gene, and can
 be used in the method of the invention. The method is for detecting
 CC prostate cancer comprising contacting a biological sample with an agent
 CC able to bind an immunogenic portion of a prostate protein (such as
 CC encoded by this sequence). An antibody which binds to an immunogenic
 CC portion of the prostate protein, and the method can be used to detect,
 CC monitor progression of, or treat prostate cancers. The antibody may
 CC also be conjugated to a therapeutic agent for use in therapy of prostate
 CC cancers.
 XX
 SQ Sequence 444 BP; 120 A; 100 C; 78 G; 143 T; 3 other;

Query Match 54.3%; Score 219; DB 19; Length 444;
 Best Local Similarity 99.6%; Pred. No. 2.le-104;
 Matches 269; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccagattcttgattgtcaaaaggatttaattgttctcagcttgggcacttcagttagg 60
 |||||
 Db 38 ccagattcttgattgtcaaaaggatttaattgttctcagcttgggcacttcagttagg 97
 |||||
 QY 61 acctaaggatgcagccgaggtttatatatgcagcaacaattattcaagcgcgacaaca 120
 |||||
 Db 98 acctaaggatgcagccgaggtttatatatgcagcaacaattattcaagcgcgacaaca 157
 |||||
 QY 121 ggatttgaacttgcgcgcagttgaatttcattccattgacttgggattcccttatc 180
 |||||
 Db 158 ggatttgaacttgcgcgcagttgaatttcattccattgacttgggattcccttatc 217
 |||||
 QY 181 agccagagagattgaaaaattaccctacgactcttactctctgtgagagggccagtggt 240
 |||||

XX AAS15797;
AC 16-JAN-2002 (first entry)
XX Human Six-Transmembrane Protein of Prostate 1, STMP1, exon 4.
XX
DE Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
XX benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ds; exon 4.
XX
OS Homo sapiens.
XX WO200172962-A2.
XX 04-OCT-2001.
XX 23-MAR-2001; 2001WO-US09410.
XX 24-MAR-2000; 2000US-191929P.
XX (SAAT/) SAATCIOGLU F.
XX Saatcioglu F;
XX WPI; 2001-662926/76.
XX
XX The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence represents exon 4 of a prostate specific protein,
CC Six-Transmembrane Protein of Prostate 1, STMP1.
XX
XX Sequence 528 BP; 139 A; 116 C; 102 G; 171 T; 0 other;
SQ
Query Match 57.8%; Score 233; DB 22; Length 528;
Best Local Similarity 100.0%; Pred. No. 1e-111;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 83 gttatatatgcagcaacaattattcaagcgcgacacagggttattgaacttgcgcgcag 142
Db 1 gttatatatgcagcaacaattattcaagcgcgacacagggttattgaacttgcgcgcag 60
Qy 143 ttgaatttcattccattgacttggatccttattcatcagccagagattgaaattta 202
Db 61 ttgaatttcattccattgacttggatccttattcatcagccagagattgaaattta 120
Qy 203 cccctagactcttactctctgagagggccagtggtggtagctataagcttgccaca 262
Db 121 cccctagactcttactctctgagagggccagtggtggtagctataagcttgccaca 180
Qy 263 ttttttttcctttactcttctgagagagtggtggttattccatccatgctagaaa 315
|||||

Db 181 ttttttctcttattctctgtcagagatgtgattccatccatgctagaaa 233
RESULT 10
AAS15806
ID AAS15806 standard; DNA; 528 BP.
XX
XX AAS15806;
XX 16-JAN-2002 (first entry)
XX Human Six-Transmembrane Protein of Prostate 1, STMP1, ORF2 exon 4.
XX
XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ds; exon 4; ORF2.
XX
XX Homo sapiens.
XX OS
XX WO200172962-A2.
XX PN
XX 04-OCT-2001.
XX PD
XX 23-MAR-2001; 2001WO-US09410.
XX PF
XX 24-MAR-2000; 2000US-191929P.
XX PR
XX (SAAT/) SAATCIOGLU F.
XX PA
XX Saatcioglu F;
XX PI
XX WPI; 2001-662926/76.
XX DR
XX
XX New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids -
XX
XX Claim 6; Fig 4F; 114pp; English.
XX
XX The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence represents exon 4 of a prostate specific protein,
CC Six-Transmembrane Protein of Prostate 1, STMP1, alternatively
CC spliced version, ORF2.
XX
XX Sequence 528 BP; 139 A; 116 C; 102 G; 171 T; 0 other;
SQ
Query Match 57.8%; Score 233; DB 22; Length 528;
Best Local Similarity 100.0%; Pred. No. 1e-111;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 83 gttatatatgcagcaacaattattcaagcgcgacacagggttattgaacttgcgcgcag 142
Db 1 gttatatatgcagcaacaattattcaagcgcgacacagggttattgaacttgcgcgcag 60
Qy 143 ttgaatttcattccattgacttggatccttattcatcagccagagattgaaattta 202
Db 61 ttgaatttcattccattgacttggatccttattcatcagccagagattgaaattta 120

XX The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence encodes prostate specific protein, Six-Transmembrane
CC Protein of Prostate 1, STMP1, ORF3.
XX
SQ Sequence 2102 BP; 595 A; 424 C; 422 G; 561 T; 0 other;

Query Match 65.5%; Score 264; DB 22; Length 2102;
Best Local Similarity 99.7%; Pred. No. 6.7e-128;
Matches 314; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccagatctttgattgtcaaggatttaattgtgtctcagcttgggcacttcagttagg 60
Db 581 ccagatctttgattgtcaaggatttaattgtgtctcagcttgggcacttcagttagg 640

Qy 61 acctaagatgcccagcggcaggtttatatatgcagcaacaattatcaagcgcgacaaca 120
Db 641 acctaagatgcccagcggcaggtttatatatgcagcaacaattatcaagcgcgacaaca 700

Qy 121 gttattgaacttgcgccaggtgaatttccattccacttgacttggatccttatcatc 180
Db 701 gttattgaacttgcgccaggtgaatttccattccacttgacttggatccttatcatc 760

Qy 181 agccagagagattgaaattaccctcagactcttactctctggagagggccagttggt 240
Db 761 agccagagagattgaaattaccctcagactcttactctctggagagggccagttggt 820

Qy 241 ggtagctataagctggccacatttttctcttattcttcttcttgcagagatgtgattca 300
Db 821 ggtagctataagctggccacatttttctcttattcttcttcttgcagagatgtgattca 880

Qy 301 tccatagctagaaa 315
Db 881 tccatagctagaaa 895

RESULT 8
AAS15801
ID AAS15801 standard; cDNA; 4329 BP.
XX
AC AAS15801;
XX
DT 16-JAN-2002 (first entry)
XX
DE Human ORF of Six-Transmembrane Protein of Prostate 1, STMP1.
XX
KW Human; Six-Transmembrane Protein of prostate 1; STMP1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 178..1650
FT /*tag= a
FT /product= "STMP1"
XX

PN WO200172962-A2.
XX
PD 04-OCT-2001.
XX
PF 23-MAR-2001; 2001WO-US09410.
XX
PR 24-MAR-2000; 2000US-191929P.
XX
PA (SAAT/) SAATCIOGLU F.
XX
PI Saatioglu F;
XX
DR WPI; 2001-662926/76.
DR P-PSDB; AAU10187.
XX
PT New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids
XX
PS Claim 5; Fig 4D; 114pp; English.
XX
CC The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence represents the open reading frame of a prostate
CC specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.
XX
SQ Sequence 4329 BP; 1315 A; 817 C; 790 G; 1407 T; 0 other;

Query Match 65.5%; Score 264; DB 22; Length 4329;
Best Local Similarity 99.7%; Pred. No. 6.8e-128;
Matches 314; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccagatctttgattgtcaaggatttaattgtgtctcagcttgggcacttcagttagg 60
Db 588 ccagatctttgattgtcaaggatttaattgtgtctcagcttgggcacttcagttagg 647

Qy 61 acctaagatgcccagcggcaggtttatatatgcagcaacaattatcaagcgcgacaaca 120
Db 648 acctaagatgcccagcggcaggtttatatatgcagcaacaattatcaagcgcgacaaca 707

Qy 121 gttattgaacttgcgccaggtgaatttccattccacttgacttggatccttatcatc 180
Db 708 gttattgaacttgcgccaggtgaatttccattccacttgacttggatccttatcatc 767

Qy 181 agccagagagattgaaattaccctcagactcttactctctggagagggccagttggt 240
Db 768 agccagagagattgaaattaccctcagactcttactctctggagagggccagttggt 827

Qy 241 ggtagctataagctggccacatttttctcttattcttcttgcagagatgtgattca 300
Db 828 ggtagctataagctggccacatttttctcttattcttcttgcagagatgtgattca 887

Qy 301 tccatagctagaaa 315
Db 888 tccatagctagaaa 902

RESULT 9
AAS15797
ID AAS15797 standard; DNA; 528 BP.

PA (UROG-) UROGENESYS INC.
XX Afar DEH, Hubert RS, Raitano AB, Saffran DC, Mitchell SC, Faris M;
PI Jakobovits A;
XX
XX
DR WPI: 2001-367804/38.
DR P-PSDB; AAE02781, AAE02841.
XX
XX
PT New STEAP (six transmembrane epithelial antigen of the prostate)
PT proteins, expressed in human cancers, useful for detecting and treating
PT cancer -
XX
XX
PS Claim 4; Fig 9A-9D; 187pp; English.
XX
XX
CC The present sequence is human six transmembrane epithelial antigen of
CC the prostate (STEAP)-2 clone GTD3 cDNA. STEAP is a member of cell
CC surface serpentine transmembrane antigens. STEAP-2 gene is located on
CC chromosome 7q21 and is used in gene therapy. Inhibiting the development
CC or progression of a cancer (eg. prostate, colon, bladder, lung, ovarian
CC and pancreatic) expressing STEAP or inhibiting growth or killing cells
CC expressing STEAP in a patient, comprises administering a vaccine
CC composition to the patient. Treating a patient with a cancer that
CC expresses STEAP, or inhibiting growth or killing cells expressing STEAP,
CC comprises administering to the patient a vector encoding single chain
CC monoclonal antibody that comprises the variable domains of the heavy and
CC light chains of the monoclonal antibody that specifically binds to STEAP,
CC such that the vector delivers the single chain monoclonal antibody coding
CC sequence to the cancer cells and the encoded single chain monoclonal
CC antibody is expressed intracellularly.
CC Note: The present sequence is also shown in sequence listing of the
CC specification, but it lacks nucleotides at its 5' end.
XX
XX
SQ Sequence 2453 BP; 648 A; 537 C; 520 G; 748 T; 0 other;

Query Match 78.2%; Score 315; DB 22; Length 2453;
Best Local Similarity 100.0%; Pred. No. 1.5e-154;
Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccagattcttgattgtcaaaaggatttaattgttctcagcttgggcacttcagttagg 60
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
765 ccagattcttgattgtcaaaaggatttaattgttctcagcttgggcacttcagttagg 824

QY 61 acctaaagatgcagccgagcaggtttatatatgcagcaacaattattcaagcgcgacaaca 120
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
825 acctaaagatgcagccgagcaggtttatatatgcagcaacaattattcaagcgcgacaaca 884

QY 121 gttattgaacttgcgcgcagttgaatttcattccattgacttgggacttccttatcatc 180
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
885 ggtattgaacttgcgcgcagttgaatttcattccattgacttgggacttccttatcatc 944

QY 181 agccagagagattgaaaaattaccctacgactcttactctcttgagagggccagtggt 240
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
945 agccagagagattgaaaaattaccctacgactcttactctcttgagagggccagtggt 1004

QY 241 ggtagctataagcttgccacatttttttcccttattcctttgttcagagatgtattca 300
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1005 ggtagctataagcttgccacatttttttcccttattcctttgttcagagatgtattca 1064

QY 301 tccatatgctagaaa 315
DB ||||||||||||||||
1065 tccatatgctagaaa 1079

RESULT 5
AAS64300
ID AAS64300 standard; cDNA; 3900 BP.
XX
AC AAS64300;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #104.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR P-PSDB; ABG00113.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 104; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3900 BP; 1161 A; 798 C; 892 G; 1049 T; 0 other;

Query Match 78.2%; Score 315; DB 23; Length 3900;
Best Local Similarity 100.0%; Pred. No. 1.5e-154;
Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccagattcttgattgtcaaaaggatttaattgttctcagcttgggcacttcagttagg 60
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
705 ccagattcttgattgtcaaaaggatttaattgttctcagcttgggcacttcagttagg 764

QY 61 acctaaagatgcagccgagcaggtttatatatgcagcaacaattattcaagcgcgacaaca 120
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
765 acctaaagatgcagccgagcaggtttatatatgcagcaacaattattcaagcgcgacaaca 824

QY 121 ggtattgaacttgcgcgcagttgaatttcattccattgacttgggacttccttatcatc 180
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
825 ggtattgaacttgcgcgcagttgaatttcattccattgacttgggacttccttatcatc 884

QY 181 agccagagagattgaaaaattaccctacgactcttactctcttgagagggccagtggt 240
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
885 agccagagagattgaaaaattaccctacgactcttactctcttgagagggccagtggt 944


```
XX
PI Saatcioglu F;
XX
DR WPI: 2001-662926/76.
DR P-PSDB; AAU10188.
XX
PT New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids.
XX
PS Claim 4; Fig 4H; 114pp; English.
XX
CC The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence encodes prostate specific protein, Six-Transmembrane
CC Protein of Prostate 1, STMP1, ORF2.
XX
SQ Sequence 1561 BP; 420 A; 319 C; 346 G; 476 T; 0 other;

Query Match 78.2%; Score 315; DB 22; Length 1561;
Best Local Similarity 100.0%; Pred. No. 1.5e-154;
Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccagattcttgattgtcaaggatttaattgttctcagcttgggcacttcagttagg 60
DB 598 cccagattcttgattgtcaaggatttaattgttctcagcttgggcacttcagttagg 657
QY 61 acctaaagatgcagccgcaggttttatatgcagcaacaattcgaagcgcaaca 120
DB 658 acctaaagatgcagccgcaggttttatatgcagcaacaattcgaagcgcaaca 717
QY 121 ggtattgaacttgcgcgcagttgaatttcattccattgacttgggtccttatc 180
DB 718 ggtattgaacttgcgcgcagttgaatttcattccattgacttgggtccttatc 777
QY 181 agccagagagattgaaatttaccctacagacttcttactctctggagagggccagtgt 240
DB 778 agccagagagattgaaatttaccctacagacttcttactctctggagagggccagtgt 837
QY 241 ggtagctataagcttgccacattttttcttattcttattcttctgacagatgtattca 300
DB 838 ggtagctataagcttgccacattttttcttattcttattcttctgacagatgtattca 897
QY 301 tccatattgctagaaa 315
DB 898 tccatattgctagaaa 912

RESULT 2
AAS76493
ID AAS76493 standard; cDNA; 2192 BP.
XX
AC AAS76493;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #12297.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
```

```
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PR
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR P-PSDB; ABG12306.
DR
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
PS Claim 1; SEQ ID No 12297; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2192 BP; 559 A; 507 C; 551 G; 575 T; 0 other;

Query Match 78.2%; Score 315; DB 23; Length 2192;
Best Local Similarity 100.0%; Pred. No. 1.5e-154;
Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccagattcttgattgtcaaggatttaattgttctcagcttgggcacttcagttagg 60
DB 705 cccagattcttgattgtcaaggatttaattgttctcagcttgggcacttcagttagg 764
QY 61 acctaaagatgcagccgcaggttttatatgcagcaacaattcgaagcgcaaca 120
DB 765 acctaaagatgcagccgcaggttttatatgcagcaacaattcgaagcgcaaca 824
QY 121 ggtattgaacttgcgcgcagttgaatttcattccattgacttgggtccttatc 180
DB 825 ggtattgaacttgcgcgcagttgaatttcattccattgacttgggtccttatc 884
QY 181 agccagagagattgaaatttaccctacagacttcttactctctggagagggccagtgt 240
DB 885 agccagagagattgaaatttaccctacagacttcttactctctggagagggccagtgt 944
QY 241 gtagctataagcttgccacattttttcttattcttattcttctgacagatgtattca 300
DB 944 gtagctataagcttgccacattttttcttattcttattcttctgacagatgtattca 300
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:59:47 ; Search time 600.75 seconds
(without alignments)
1151.755 Million cell updates/sec

Title: US-09-802-520-5
Perfect score: 403

Sequence: 1 cccgattcttctgattgtca.....cttgctccctccctagatatac 403

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/hold-genseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	315	78.2	1561	22 AAS15811	Human cDNA encoding
2	315	78.2	2192	23 AAS76493	DNA encoding novel
3	315	78.2	2238	22 AAS15810	Human ORF2 of Six-
4	315	78.2	2453	22 AAD07072	Human six transmem
5	315	78.2	3900	23 AAS64300	DNA encoding novel
6	264	65.5	1680	22 AAS15802	Human cDNA encoding
7	264	65.5	2102	22 AAS15813	Human cDNA encoding
8	264	65.5	4329	22 AAS15801	Human ORF of Six-T
9	233	57.8	528	22 AAS15797	Human Six-Transmem

10	233	57.8	528	22	AAS15806	Human Six-Transmem
11	219	54.3	444	19	AAV61348	Forward DNA sequen
12	219	54.3	444	19	AAV58684	Prostate tumour sp
13	219	54.3	444	21	AAO6447	Human immunogenic
14	219	54.3	444	22	AAS63655	Human prostate CDN
15	219	54.3	444	22	AAS10206	Human prostate tum
16	219	54.3	444	22	AAH93363	Human prostate-spe
17	219	54.3	444	22	AAH84877	Human prostate-spe
18	219	54.3	444	22	AAH02628	Prostate tumour an
19	185	45.9	1725	22	AAS15793	Human DNA for Six-
20	153	38.0	328	19	AAV61346	Forward DNA sequen
21	153	38.0	328	19	AAV58682	Prostate tumour sp
22	153	38.0	328	21	AAO6445	Human immunogenic
23	153	38.0	328	22	AAS63653	Human prostate CDN
24	153	38.0	328	22	AAS10204	Human prostate tum
25	153	38.0	328	22	AAH93361	Human prostate-spe
26	153	38.0	328	22	AAH84875	Human prostate-spe
27	153	38.0	328	22	AAH02626	Prostate tumour an
28	99	24.6	395	22	ABA57377	Human foetal liver
29	99	24.6	395	22	ABA26909	Probe #5375 for ge
30	99	24.6	395	22	AAK05412	Human brain expres
31	99	24.6	395	22	AAK31011	Human bone marrow
32	99	24.6	395	22	AAI15504	Probe #5437 for ge
33	99	24.6	395	22	AAI36922	Probe #5608 used t
34	84	20.8	539	22	ABA50938	Human breast cell
35	84	20.8	539	22	ABA68917	Human foetal liver
36	84	20.8	539	22	ABA35870	Human brain expres
37	84	20.8	539	22	AAK17249	Human bone marrow
38	84	20.8	539	22	AAK43038	Probe #13734 for g
39	84	20.8	539	22	AAI23801	Probe #17799 used
40	84	20.8	539	22	AAI49113	Probe #9401 used t
41	84	20.8	539	22	AAI09410	Human breast cell
42	84	20.8	1953	22	ABA45820	Human foetal liver
43	84	20.8	1953	22	ABA56335	Probe #4442 for ge
44	84	20.8	1953	22	ABA25976	Human brain expres
45	84	20.8	1953	22	AAK04514	

ALIGNMENTS

RESULT 1

AAS15811
ID AAS15811 standard; cDNA; 1561 BP.

XX AC AAS15811;

XX DT 16-JAN-2002 (first entry)

XX Human cDNA encoding ORF2 of Six-Transmembrane Protein of Prostate 1.
XX DE Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
XX DE benign prostatic hyperplasia; acute prostatitis; testicular cancer;
XX DE cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
XX DE leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
XX DE liver cancer; lung cancer; cytostatic; ss; ORF2.

XX OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 188..1552

FT /*tag= a

FT /product= "STMP1, ORF2"

XX WO200172962-A2.

XX PD 04-OCT-2001.

XX PF 23-MAR-2001; 2001WO-0509410.

XX PR 24-MAR-2000; 2000US-191929P.

XX PA (SAAT/) SAATCIOGLU F.

THIS PAGE BLANK (USPTO)

McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y.,
Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
Tsurgon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 205085)
Green, E.D.
Direct Submission
Submitted (24-MAR-2000) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
On Oct 27, 2001 this sequence version replaced gi:7321458.

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc_mouse@nhgri.nih.gov

----- Project Information

Center project name: xh

Center clone name: 119M19

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 193516 bases at least Q40

Consensus quality: 200596 bases at least Q30

Consensus quality: 201000 bases at least Q20

Insert size: 197000; agarose-fp

Insert size: 192000; pulse-field-gel

Insert size: 204185; sum-of-contigs

Quality coverage: 9.36x in Q20 bases; agarose-fp

Quality coverage: 9.60x in Q20 bases; pulse-field-gel

Quality coverage: 9.03x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 4428: contig of 4428 bp in length
* 4429 4528: gap of unknown length
* 4529 9553: contig of 5025 bp in length
* 9554 9654: gap of unknown length
* 9654 22746: contig of 13093 bp in length
* 22747 22846: gap of unknown length
* 22847 33954: contig of 11108 bp in length
* 33955 34055: gap of unknown length
* 34055 47831: contig of 13777 bp in length
* 47832 47931: gap of unknown length
* 47932 62010: contig of 14079 bp in length
* 62011 62110: gap of unknown length
* 62111 81789: contig of 19679 bp in length
* 81790 81889: gap of unknown length
* 81890 109919: contig of 28030 bp in length
* 109920 110019: gap of unknown length
* 110020 153235: contig of 43216 bp in length
* 153236 153335: gap of unknown length
* 153336 205085: contig of 51750 bp in length.

FEATURES

source

1..205085
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP23-119M19"
/clone_lib="RPCI mouse BAC library 23"
1..4428
/note="assembly_fragment"
4529..9553
/note="assembly_fragment"

misc_feature

misc_feature

misc_feature 9654..22746
/note="assembly_fragment"
misc_feature 22847..33954
/note="assembly_fragment"
misc_feature 34055..47831
/note="assembly_fragment"
clone_end:77
vector_side:right
misc_feature 47932..62010
/note="assembly_fragment"
misc_feature 62111..81789
/note="assembly_fragment"
misc_feature 81890..109919
/note="assembly_fragment"
misc_feature 110020..153235
/note="assembly_fragment"
misc_feature 153336..205085
/note="assembly_fragment"
clone_end:SP6
vector_side:left
BASE COUNT 61131 a 40716 c 41296 g 61026 t 916 others
ORIGIN

Query Match 6.5%; Score 26; DB 2; Length 205085;
Best Local Similarity 100.0%; Pred. No. 0.0048;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 caggtttatatgcagcaacaatat 105
|||||

Db 191831 CAGGTTTATATGCGACCAACAATAT 191856

Search completed: September 20, 2002, 09:53:50
Job time: 13088 sec

```

ACCESSION AX200582
VERSION AX200582.1 GI:15390402
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 328)
Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
Reed, S.G., Kalos, M.D., Fanger, G.R., Day, C.H., Retter, M.W.,
Stolk, J.A., Skeiky, Y.A., Wang, A. and Meagher, M.J.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0151633-A 212 19-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..328
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 84 a 73 c 61 g 106 t 4 others
ORIGIN

Query Match 38.0%; Score 153; DB 6; Length 328;
Best Local Similarity 99.5%; Pred. No. 6.1e-72;
Matches 203; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 attcttgattgtcaaggatttaattgtctcagcttggtgagcttcagtttagaccta 65
Db 44 ATTCTTTGATTGTCAAGGATTTAATGTGTCTCAGCTTGGGCACTTCAGTTAGACCTA 103
Qy 66 aggatgccagccgaggtttatatatgcagcaacaattccaagcgcgacaacaggtta 125
Db 104 AGGATGCCAGCCGCGAGGTTTATATATGCAGCAACAATATTCAGCGGCAACAGGTTA 163
Qy 126 ttgaactggccgcagtggaatttcattccattgactgggtgccttcatcagcca 185
Db 164 TTGAAC TTGCCCGCAGTTNAATTTCATCCATTGACTTGGGATCCTTATCATCAGCCA 223
Qy 186 gagagattgaaatttaccctac 209
Db 224 GAGAGATTGAAATTTACCCCTAC 247

RESULT 13
AX267238
LOCUS AX267238 328 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 212 from Patent WO0173032.
ACCESSION AX267238
VERSION AX267238.1 GI:16516016
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
Kalos, M.D., Fanger, G.R., Retter, M.W., Stolk, J.A., Day, C.H.,
Vedvick, T.S., Carter, D., Li, S.X., Wang, A., Skeiky, Y.A., Hepler, W.T.
and Henderson, R.A.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0173032-A 212 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..328
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 84 a 73 c 61 g 106 t 4 others
ORIGIN

Query Match 38.0%; Score 153; DB 6; Length 328;

```

```

Best Local Similarity 99.5%; Pred. No. 6.1e-72;
Matches 203; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 attcttgattgtcaaggatttaattgtctcagcttggtgagcttcagtttagaccta 65
Db 44 ATTCTTTGATTGTCAAGGATTTAATGTGTCTCAGCTTGGGCACTTCAGTTAGACCTA 103
Qy 66 aggatgccagccgaggtttatatatgcagcaacaattccaagcgcgacaacaggtta 125
Db 104 AGGATGCCAGCCGCGAGGTTTATATATGCAGCAACAATATTCAGCGGCAACAGGTTA 163
Qy 126 ttgaactggccgcagtggaatttcattccattgactgggtgccttcatcagcca 185
Db 164 TTGAAC TTGCCCGCAGTTNAATTTCATCCATTGACTTGGGATCCTTATCATCAGCCA 223
Qy 186 gagagattgaaatttaccctac 209
Db 224 GAGAGATTGAAATTTACCCCTAC 247

RESULT 14
AX155253
LOCUS AX155253 322 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 11 from Patent WO0140276.
ACCESSION AX155253
VERSION AX155253.1 GI:14536717
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 322)
Afar, D.E., Hubert, R.S., Raitano, A.B., Saffran, D.C., Mitchell, S.C.,
Paris, M. and Jakobovits, A.
TITLE Serpentine transmembrane antigens expressed in human prostate
cancer and uses thereof
JOURNAL Patent: WO 0140276-A 11 07-JUN-2001;
Urogenesys, Inc. (US)
FEATURES
source Location/Qualifiers
1..322
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 84 a 70 c 62 g 106 t
ORIGIN

Query Match 12.4%; Score 50; DB 6; Length 322;
Best Local Similarity 100.0%; Pred. No. 6.3e-16;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 342 ctatagagattgtgaataaaaccttaacctatagttgacctacttgc 391
Db 83 CTATAGAGATTGTGAATAAAACCTTACCTATAGTTGCCATTACTTGTCTC 132

RESULT 15
AC026813
LOCUS AC026813 205085 bp DNA linear HTG 27-OCT-2001
DEFINITION Mus musculus chromosome 5 clone RP23-119M19 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 10 unordered pieces.
ACCESSION AC026813
VERSION AC026813.2 GI:16506398
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 205085)
Akhter, N., Avele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-tin, S.-Q.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Masiello, C., Mastrian, S.D.,

```



```

* 1 2438: contig of 2438 bp in length
* 2439 2538: gap of unknown length
* 2539 8133: contig of 5595 bp in length
* 8134 8233: gap of unknown length
* 8234 40378: contig of 32145 bp in length
* 40379 40478: gap of unknown length
* 40479 172915: contig of 132437 bp in length.
FEATURES
    source
        1..172915
            /organism="Papio cynocephalus anubis"
            /db_xref="taxon:9555"
            /clone="RP41-167P22"
            /clone_lib="RP41"
        1..2438
            /note="assembly_fragment"
            clone_end:SP6
            vector_side:right"
        2539..8133
            /note="assembly_fragment"
        8234..40378
            /note="assembly_fragment"
            clone_end:T7
            vector_side:right"
        40479..172915
            /note="assembly_fragment"

```

```

BASE COUNT 52509 a 31733 c 32277 g 56096 t 300 others
ORIGIN
Query Match 39.0%; Score 157; DB 2; Length 172915;
Best Local Similarity 100.0%; Pred. No. 2.8e-74;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 159 ttgacttgggaccccttatcatcagccagagagattgaaatttaccctacagactcttta 218
|||||
Db 85956 TTGACTTGGATCCTTATCATCAGCCAGAGAGATTGAAATTTACCCCTACGACTTTTA 86015
QY 219 ctctctgagagggccagtggtgtagctataagcttgccacattttttcccttatt 278
|||||
Db 86016 CTCTCTGAGAGGGCCAGTGGTGGTAGCTATAGCTTGGCCACATTTTTTTCCTTTATT 86075
QY 279 ctttgtcagagatgtgattccatcctatgctagaaa 315
Db 86076 CCTTTGTGACAGATGTGATTCATCCATATGCTAGAAA 86112

```

```

RESULT 10
AX106431
LOCUS AX106431 328 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 212 from Patent WO0125272.
ACCESSION AX106431
VERSION AX106431.1 GI:13922110
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Xu,J., Skelky,Y.A., Reed,S.G. and Cheever,M.A.
TITLE Compositions and methods for therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0125272-A 212 12-APR-2001;
CORIXA CORPORATION (US)
FEATURES
    source
        1..328
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
        1..328
            /note="n = A,T,C or G"
    misc_feature 84 a 73 c 61 g 106 t 4 others
BASE COUNT 84 a 73 c 61 g 106 t 4 others
ORIGIN

```

```

Query Match 38.0%; Score 153; DB 6; Length 328;
Best Local Similarity 99.5%; Pred. No. 6.1e-72;
Matches 203; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 6 attcttggattgtcaagagatttaattgtctcagcttggcacttcagttaggacctta 65
|||||
Db 44 ATTCITTTGATTGTCAAGAGATTAAATGTGTCTCAGCTTGGGCACCTTCAGTTAGGACCTA 103
QY 66 aggtatgccagccggcagggtttatatatgcagcaacaatttcaagcggacacagggtta 125
|||||
Db 104 AGGATGCCAGCCGGCAGGTTTATATATGCAGCAACAATATTCAAGCGGCACACAGGTTA 163
QY 126 ttgaacttgcgcgcaggttgaatttcattcccatgacttggatgaccttatcatcagcca 185
|||||
Db 164 TTGAACCTTGCCTCCGAGTTTAAATTTCCATTCCTCCATTTGGGATCCTTATCATCAGCCA 223
QY 186 gagagattgaaaatttacccttac 209
|||||
Db 224 GAGAGATTGAAAATTTACCCCTAC 247

```

```

RESULT 11
AX140722
LOCUS AX140722 328 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 212 from Patent WO0134802.
ACCESSION AX140722
VERSION AX140722.1 GI:14280833
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,
Skelky,Y.A. and Wang,A.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0134802-A 212 17-MAY-2001;
CORIXA CORPORATION (US)
FEATURES
    source
        1..328
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
        84 a 73 c 61 g 106 t 4 others
BASE COUNT 84 a 73 c 61 g 106 t 4 others
ORIGIN

```

```

Query Match 38.0%; Score 153; DB 6; Length 328;
Best Local Similarity 99.5%; Pred. No. 6.1e-72;
Matches 203; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 6 attcttggattgtcaagagatttaattgtctcagcttggcacttcagttaggacctta 65
|||||
Db 44 ATTCITTTGATTGTCAAGAGATTAAATGTGTCTCAGCTTGGGCACCTTCAGTTAGGACCTA 103
QY 66 aggtatgccagccggcagggtttatatatgcagcaacaatttcaagcggacacagggtta 125
|||||
Db 104 AGGATGCCAGCCGGCAGGTTTATATATGCAGCAACAATATTCAAGCGGCACACAGGTTA 163
QY 126 ttgaacttgcgcgcaggttgaatttcattcccatgacttggatgaccttatcatcagcca 185
|||||
Db 164 TTGAACCTTGCCTCCGAGTTTAAATTTCCATTCCTCCATTTGGGATCCTTATCATCAGCCA 223
QY 186 gagagattgaaaatttacccttac 209
|||||
Db 224 GAGAGATTGAAAATTTACCCCTAC 247

```

```

RESULT 12
AX200582
LOCUS AX200582 328 bp DNA linear PAT 29-AUG-2001
DEFINITION Sequence 212 from Patent WO0151633.

```

JOURNAL

Submitted (12-DEC-2001) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA

COMMENT

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc.mouse@nhgri.nih.gov

----- Project Information

Center project name: ces

Center clone name: 120K11

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 161001 bases at least Q40

Consensus quality: 161422 bases at least Q30

Consensus quality: 161562 bases at least Q20

Insert size: 143000; agarose-fp

Insert size: 162428; sum-of-contigs

Quality coverage: 10.76x in Q20 bases; agarose-fp

Quality coverage: 9.47x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 13922: contig of 13922 bp in length
* 13923 14022: gap of unknown length
* 14023 36248: contig of 22226 bp in length
* 36249 36348: gap of unknown length
* 36349 66192: contig of 29844 bp in length
* 66193 66292: gap of unknown length
* 66293 92168: contig of 25876 bp in length
* 92169 92268: gap of unknown length
* 92269 122037: contig of 29768 bp in length
* 122037 122136: gap of unknown length
* 122137 162928: contig of 40792 bp in length.

FEATURES

source

1. 162928
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-120K11"
/clone_lib="RP43"

misc_feature

1. 13922
/note="assembly_fragment"
clone_end:T7
vector_side:left

misc_feature

14023..36248

misc_feature

/note="assembly_fragment"

misc_feature

36349..66192

misc_feature

/note="assembly_fragment"

misc_feature

66293..92168

misc_feature

/note="assembly_fragment"

misc_feature

92269..122036

misc_feature

/note="assembly_fragment"

misc_feature

122137..162928

misc_feature

/note="assembly_fragment"

misc_feature

vector_side:right

BASE COUNT 51528 a 29878 c 30533 g 50485 t 504 others
ORIGIN

Query Match 50.4%; Score 203; DB 2; Length 162928;
Best Local Similarity 100.0%; Pred. No. 2,7e-99;
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 113 cgacaacaggttattgaacttgcgcagtgcaatttcattccattgacttgggatcc 172
|||||

Db 121770 CGACAACAGGTTATTGAACCTTGCCTCCGAGTTGAATTTTCATTCATTCAGTGGATCC 121711
Qy 173 ttatcatcagccagagagattgaaattacccttagactcttactctctctgagagagg 232
|||||
Db 121710 TTATCATCAGCCAGAGAGATTGAAAATTTACCCCTAGACACTTTACTCTCTGGAGAGGG 121651
Qy 233 ccagtgtggtagctataagcttggccacatttttttcccttttcttcttcttcttctgagagat 292
|||||
Db 121650 CCACTGTGGTAGCTATAAGCTTGGCCACATTTTTTTTCCCTTTATTCCTTTGTGAGAGAT 121591
Qy 293 gtgattcatccatagctagaaa 315
|||||
Db 121590 GTGATTTCATCCATATGCTAGAAA 121568

RESULT 9

AC099742

LOCUS

DEFINITION Papi0 cynocephalus anubis clone RP41-167P22, WORKING DRAFT
AC099742 172915 bp DNA linear HTG 20-NOV-2001
SEQUENCE, 4 unordered pieces.

AC099742

AC099742.1

GI:17017546

KEYWORDS

HTG; HTGS-PHASE1; HTGS-DRAFT.

SOURCE

olive baboon.

ORGANISM

Papi0 cynocephalus anubis

REFERENCE

AUTHORS

1 (bases 1 to 172915)
Akhtar, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-Q.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Masiello, C., Mastrian, S.D.,
McCluskey, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, I.,
Stantropop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc.mouse@nhgri.nih.gov
----- Project Information
Center project name: ccv
Center clone name: 167P22
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171178 bases at least Q40
Consensus quality: 171910 bases at least Q30
Consensus quality: 172236 bases at least Q20
Insert size: 130000; agarose-fp
Insert size: 172615; sum-of-contigs
Quality coverage: 12.58x in Q20 bases; agarose-fp
Quality coverage: 9.47x in Q20 bases; sum-of-contigs
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

QY 121 ggttattgaactgcccgcagtgtaattcattccattgacttggtggtccttattc 180
 |||||
 Db 158 GGTATTGAACTTGC CGCAGTTGAATTTCATTCCTCCATTGACTTGGGATCCTTATCATC 217
 |||||
 QY 181 acccagagattgaaattaccctacagactcttactctctgagagggccagtgg 240
 |||||
 Db 218 ACCANAGAGATTGAAATTTACCCCTACGACTCTTACTCTCTGGAGGGCCAGTGGT 277
 |||||
 QY 241 ggtagctataagcttgccacattttttt 270
 |||||
 Db 278 GGTAGCTATAAGCTTGCCACATTTTTTTT 307
 |||||

RESULT 6
 AX200584
 LOCUS AX200584 444 bp DNA linear PAT 29-AUG-2001
 DEFINITION Sequence 214 from Patent WO0151633.
 ACCESSION AX200584
 VERSION AX200584.1 GI:15390404
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 444)
 AUTHORS Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
 Reed, S.G., Kalos, M.D., Fanger, G.R., Retter, M.W.,
 Stolk, J.A., Skeiky, Y.A., Wang, A., and Meagher, M.J.
 TITLE Compositions and methods for the therapy and diagnosis of prostate
 cancer
 JOURNAL Patent: WO 0151633-A 214 19-JUL-2001;
 CORIXA CORPORATION (US)
 FEATURES
 source Location/Qualifiers
 i. .444
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 120 a 100 c 78 g 143 t 3 others
 ORIGIN
 Query Match 54.3%; Score 219; DB 6; Length 444;
 Best Local Similarity 99.6%; Pred. No. 7.6e-108;
 Matches 269; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ccagattcttgattgtcaaggatttaattgtctcagcttggtggcacttcagttagg 60
 |||||
 Db 38 CCAGATTCTTTGATTGTCAAGGATTAAATGTTGTCTCAGCTTGGGCACCTTCAGTTAGG 97
 |||||
 QY 61 acctaaaggatgccagccgcaggtttatatatgcagcaacaattattcaagcgcgacaaca 120
 |||||
 Db 98 ACCTAAGGATGCCAGCGCAGGTTTATATATGCACACAATATTCAAGCGCGACAACA 157
 |||||
 QY 121 ggtattgaactgcccgcagtgtaattcattccattgacttggtggtccttattc 180
 |||||
 Db 158 GGTATTGAACTTGC CGCAGTTGAATTTCATTCCTCCATTGACTTGGGATCCTTATCATC 217
 |||||
 QY 181 agccagagattgaaattaccctacagactcttactctctgagagggccagtgg 240
 |||||
 Db 218 AGCCANAGAGATTGAAATTTACCCCTACGACTCTTACTCTCTGGAGGGCCAGTGGT 277
 |||||
 QY 241 ggtagctataagcttgccacattttttt 270
 |||||
 Db 278 GGTAGCTATAAGCTTGCCACATTTTTTTT 307
 |||||

Query Match 54.3%; Score 219; DB 6; Length 444;
 Best Local Similarity 99.6%; Pred. No. 7.6e-108;
 Matches 269; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ccagattcttgattgtcaaggatttaattgtctcagcttggtggcacttcagttagg 60
 |||||
 Db 38 CCAGATTCTTTGATTGTCAAGGATTAAATGTTGTCTCAGCTTGGGCACCTTCAGTTAGG 97
 |||||
 QY 61 acctaaaggatgccagccgcaggtttatatatgcagcaacaattattcaagcgcgacaaca 120
 |||||
 Db 98 ACCTAAGGATGCCAGCGCAGGTTTATATATGCACACAATATTCAAGCGCGACAACA 157
 |||||
 QY 121 ggtattgaactgcccgcagtgtaattcattccattgacttggtggtccttattc 180
 |||||
 Db 158 GGTATTGAACTTGC CGCAGTTGAATTTCATTCCTCCATTGACTTGGGATCCTTATCATC 217
 |||||
 QY 181 agccagagattgaaattaccctacagactcttactctctgagagggccagtgg 240
 |||||
 Db 218 AGCCANAGAGATTGAAATTTACCCCTACGACTCTTACTCTCTGGAGGGCCAGTGGT 277
 |||||
 QY 241 ggtagctataagcttgccacattttttt 270
 |||||
 Db 278 GGTAGCTATAAGCTTGCCACATTTTTTTT 307
 |||||

RESULT 7
 AX267240
 LOCUS AX267240 444 bp DNA linear PAT 26-OCT-2001
 DEFINITION Sequence 214 from Patent WO0173032.
 ACCESSION AX267240
 VERSION AX267240.1 GI:16516018
 KEYWORDS

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Xu, J., Dillon, D.C., Mitcham, J.L., Harlocker, S.L., Jiang, Y.,
 Kalos, M.D., Fanger, G.R., Retter, M.W., Stolk, J.A., Day, C.H.,
 Wedvick, T.S., Carter, D., Li, S.X., Wang, A., Skeiky, Y.A., Hepler, W.T.
 and Henderson, R.A.
 TITLE Compositions and methods for the therapy and diagnosis of prostate
 cancer
 JOURNAL Patent: WO 0173032-A 214 04-OCT-2001;
 CORIXA CORPORATION (US)
 FEATURES
 source Location/Qualifiers
 i. .444
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 120 a 100 c 78 g 143 t 3 others
 ORIGIN
 Query Match 54.3%; Score 219; DB 6; Length 444;
 Best Local Similarity 99.6%; Pred. No. 7.6e-108;
 Matches 269; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ccagattcttgattgtcaaggatttaattgtctcagcttggtggcacttcagttagg 60
 |||||
 Db 38 CCAGATTCTTTGATTGTCAAGGATTAAATGTTGTCTCAGCTTGGGCACCTTCAGTTAGG 97
 |||||
 QY 61 acctaaaggatgccagccgcaggtttatatatgcagcaacaattattcaagcgcgacaaca 120
 |||||
 Db 98 ACCTAAGGATGCCAGCGCAGGTTTATATATGCACACAATATTCAAGCGCGACAACA 157
 |||||
 QY 121 ggtattgaactgcccgcagtgtaattcattccattgacttggtggtccttattc 180
 |||||
 Db 158 GGTATTGAACTTGC CGCAGTTGAATTTCATTCCTCCATTGACTTGGGATCCTTATCATC 217
 |||||
 QY 181 agccagagattgaaattaccctacagactcttactctctgagagggccagtgg 240
 |||||
 Db 218 AGCCANAGAGATTGAAATTTACCCCTACGACTCTTACTCTCTGGAGGGCCAGTGGT 277
 |||||
 QY 241 ggtagctataagcttgccacattttttt 270
 |||||
 Db 278 GGTAGCTATAAGCTTGCCACATTTTTTTT 307
 |||||

RESULT 8
 AC104475/c
 LOCUS AC104475 162928 bp DNA linear HTG 12-DEC-2001
 DEFINITION Pan troglodytes clone RP43-120K11, WORKING DRAFT SEQUENCE, 6
 unordered pieces.
 AC104475
 AC104475.1 GI:17530717
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS chimpanzee.
 SOURCE Pan troglodytes
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 REFERENCE 1 (bases 1 to 162928)
 AUTHORS Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-Q.,
 Legaspi, R., Maduro, V.B., Masiello, C., Masiello, C., Mastrian, S.D.,
 McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Stantripop, S.,
 Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurgon, C., Vogt, J.L.,
 Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and
 Green, E.D.
 TITLE NISC Comparative Sequencing Initiative
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 162928)
 AUTHORS Green, E.D.
 TITLE Direct Submission

<http://genome.wustl.edu/gsc>
e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/STB/CH7> or send an E-mail to egreen@nhgri.nih.gov

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).
VECTOR: pBEO
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of H_RG016J04; actual end is at 156214 of H_RG016J04. The orientation of this clone is unknown.

This clone contains STS SWS52784 (NID:gil113580) and SWS5893 (NID:g454733).

FEATURES	source
	location/Qualifiers
	1..156214
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7q21"
	/clone="RG016J04"
	/clone_lib="CITB-978SK-B"
repeat_region	15..40
	/rpt_family="L1"
repeat_region	326..621
	/rpt_family="ALU"
repeat_region	complement(977..1499)
	/rpt_family="L1"
repeat_region	3398..3421
	/rpt_family="L1"
repeat_region	5319..5345
	/rpt_family="L1"
repeat_region	complement(8071..8347)
	/rpt_family="ALU"
repeat_region	complement(9406..9975)
	/rpt_family="L1"
repeat_region	complement(10000..11285)
	/rpt_family="L1"
repeat_region	complement(11315..11984)
	/rpt_family="L1"
repeat_region	11666..11687
	/rpt_family="L1"
repeat_region	complement(12010..12299)
	/rpt_family="ALU"
repeat_region	complement(12301..13893)
	/rpt_family="L1"
repeat_region	13897..14096
	/rpt_family="L1"
repeat_region	complement(14777..14838)

misc_feature	/rpt_family="L1"
	15715..15767
	/note="similar to human EST T02878 (NID:g314119)"
misc_feature	19436..19497
	/note="similar to human EST AA123941 (NID:g1682616)
repeat_region	mq22f09.rl
	complement(25113..25148)
	/rpt_family="L1"
repeat_region	complement(25561..25578)
	/rpt_family="L1"
repeat_region	complement(26832..27124)
	/rpt_family="ALU"
misc_feature	27125..27383
	/note="similar to human EST AA151807 (NID:gil720502)
	z139g02.rl
misc_feature	27125..27544
	/note="similar to human EST AA151796 (NID:gil720491)
	z139c02.rl
misc_feature	complement(27178..27600)
	/note="similar to human EST AA149579 (NID:gil720380)
	z139c02.s1
misc_feature	complement(27441..27890)
	/note="similar to human EST N52554 (NID:gil93720)
	yv36a11.s1
misc_feature	27686..27856
	/note="similar to human EST W32120 (NID:gil313113)
	zb97c08.rl
misc_feature	27686..27838
	/note="similar to human EST W32085 (NID:gil313105)
	zb97b08.rl
misc_feature	complement(27733..28143)
	/note="similar to human EST N59831 (NID:gil203721)
	yz77a10.s1
misc_feature	complement(27838..28143)
	/note="similar to human EST W31561 (NID:gil312680)
	zb97b08.s1
misc_feature	complement(27887..28143)
	/note="similar to human EST N59830 (NID:gil203720)
	yz77a09.s1
misc_feature	complement(27997..28144)
	/note="similar to human EST W31628 (NID:gil312688)
	zb97c08.s1
misc_feature	28666..28972
	/note="similar to human EST R09339 (NID:g761262)
	yf26d05.rl
misc_feature	complement(29587..29853)
	/note="similar to human EST R09227 (NID:g761150)
	yf26d05.s1
misc_feature	30146..30314
	/note="similar to human EST N78042 (NID:gil240743)
	yv71e06.rl
misc_feature	complement(31170..31350)
	/note="similar to human EST N58451 (NID:gil202341)
	yv71e06.s1
repeat_region	31970..32006
	/rpt_family="L1"
repeat_region	32263..32563
	/rpt_family="ALU"
repeat_region	complement(34737..34778)
	/rpt_family="L1"
repeat_region	36558..36709
	/rpt_family="ALU"
repeat_region	complement(37630..37740)
	/rpt_family="PTR"
repeat_region	39839..40125
	/rpt_family="ALU"
repeat_region	40126..40157
	/rpt_family="L1"
repeat_region	complement(41340..41706)
	/rpt_family="ALU"
repeat_region	complement(41904..41933)
	/rpt_family="L1"
repeat_region	42348..42409

* 19907 20006: gap of 100 bp
* 20007 20922: contig of 916 bp in length
* 20923 21022: gap of 100 bp
* 21023 21936: contig of 914 bp in length
* 21937 22036: gap of 100 bp
* 22037 22942: contig of 906 bp in length
* 22943 23042: gap of 100 bp
* 23043 23928: contig of 886 bp in length
* 23929 24028: gap of 100 bp
* 24029 24911: contig of 883 bp in length
* 24912 25011: gap of 100 bp
* 25012 25887: contig of 876 bp in length
* 25888 25987: gap of 100 bp
* 25988 26883: contig of 896 bp in length
* 26884 26983: gap of 100 bp
* 26984 27912: contig of 929 bp in length
* 27913 28012: gap of 100 bp
* 28013 28919: contig of 907 bp in length
* 28920 29019: gap of 100 bp
* 29020 29933: contig of 914 bp in length
* 29934 30033: gap of 100 bp
* 30034 30930: contig of 897 bp in length
* 30931 31030: gap of 100 bp
* 31031 31924: contig of 894 bp in length
* 31925 32024: gap of 100 bp
* 32025 32902: contig of 878 bp in length
* 32903 33002: gap of 100 bp
* 33003 33890: contig of 888 bp in length
* 33891 33990: gap of 100 bp
* 33991 34900: contig of 910 bp in length
* 34901 35000: gap of 100 bp
* 35001 35895: contig of 895 bp in length
* 35896 35995: gap of 100 bp
* 35996 36840: contig of 845 bp in length
* 36841 36940: gap of 100 bp
* 36941 37822: contig of 882 bp in length
* 37823 37922: gap of 100 bp
* 37923 38845: contig of 923 bp in length
* 38846 38945: gap of 100 bp
* 38946 39823: contig of 878 bp in length
* 39824 39923: gap of 100 bp
* 39924 40792: contig of 869 bp in length
* 40793 40892: gap of 100 bp
* 40893 41811: contig of 919 bp in length
* 41812 41911: gap of 100 bp
* 41912 42806: contig of 895 bp in length
* 42807 42906: gap of 100 bp
* 42907 43784: contig of 878 bp in length
* 43785 43884: gap of 100 bp
* 43885 44777: contig of 893 bp in length
* 44778 44877: gap of 100 bp
* 44878 45768: contig of 891 bp in length
* 45769 45868: gap of 100 bp
* 45869 46761: contig of 893 bp in length
* 46762 46861: gap of 100 bp
* 46862 47713: contig of 852 bp in length
* 47714 47813: gap of 100 bp
* 47814 48729: contig of 916 bp in length
* 48730 48823: gap of 100 bp
* 48830 49779: contig of 950 bp in length
* 49780 49879: gap of 100 bp
* 49880 50794: contig of 915 bp in length
* 50795 50894: gap of 100 bp
* 50895 51797: contig of 903 bp in length
* 51798 51897: gap of 100 bp
* 51898 52786: contig of 889 bp in length
* 52787 52886: gap of 100 bp
* 52887 53779: contig of 893 bp in length
* 53780 53879: gap of 100 bp
* 53880 54782: contig of 903 bp in length
* 54783 54882: gap of 100 bp
* 54883 55769: contig of 887 bp in length
* 55770 55869: gap of 100 bp

* 55870 56736: contig of 867 bp in length
* 56737 56836: gap of 100 bp
* 56837 57715: contig of 879 bp in length
* 57716 57815: gap of 100 bp
* 57816 58712: contig of 897 bp in length
* 58713 58812: gap of 100 bp
* 58813 59729: contig of 917 bp in length
* 59730 59829: gap of 100 bp
* 59830 60638: contig of 809 bp in length
* 60639 60738: gap of 100 bp
* 60739 61654: contig of 916 bp in length
* 61655 61754: gap of 100 bp
* 61755 62680: contig of 926 bp in length
* 62681 62780: gap of 100 bp
* 62781 63664: contig of 884 bp in length
* 63665 63764: gap of 100 bp
* 63765 64670: contig of 906 bp in length
* 64671 64770: gap of 100 bp
* 64771 65666: contig of 896 bp in length
* 65667 65766: gap of 100 bp
* 65667 66636: contig of 870 bp in length
* 66637 66736: gap of 100 bp
* 66737 67604: contig of 888 bp in length
* 67605 67704: gap of 100 bp
* 67705 68576: contig of 872 bp in length
* 68577 68676: gap of 100 bp
* 68677 69601: contig of 925 bp in length
* 69602 69701: gap of 100 bp
* 69702 70618: contig of 917 bp in length

Query Match 58.6%; Score 236; DB 2; Length 87401;
Best Local Similarity 100.0%; Pred. No. 3.le-117;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 80 caggtttatatgcagcaacaataattcaagcgcgcacacacaggttatgaacttgcgcgc 139
|||||
Db 46422 CAGGTTTATATATGCGACCAATAATTCAAGCGCGACACAGGTTATTGAATTCGCCGC 46481
|||||

Qy 140 cagttgaatttcattccattgacttggtatccttatcatcagccagagagattgaaaaa 199
|||||
Db 46482 CAGTTGAATTTTCATTCCTCCATTGCGATCCTTATCATCAGCCAGAGATTGAAAT 46541
|||||

Qy 200 ttaccctacgactctttactctctgtgagagggccagtggtagctataagcttgccc 259
|||||
Db 46542 TTACCCCTACGACTCTTTACTCTCTGAGAGGGCCAGTGGTGTAGCTATAAGCTTGGCC 46601
|||||

Qy 260 acatttttttccctttattcctttgtcagagagatgattccatcctatgctagaaa 315
|||||
Db 46602 ACATTTTTCCTTTTCTTCTTTCATCAGAGATGTGATTTCATCATATCTAGAAA 46657
|||||

RESULT 3

HSAC002064
LOCUS
DEFINITION Human BAC clone RG016J04 from 7q21, complete sequence.
ACCESSION AC002064
VERSION AC002064.1 GI:2076723
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 156214)
Gattung, S. and Maggi, L.
The sequence of H. sapiens BAC clone RG016J04
Unpublished (1997)
JOURNAL
REFERENCE 2 (bases 1 to 156214)
WATERSTON, R.
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1997)
COMMENT Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA

/note="unnamed protein product"
/codon_start=1
/protein_id="CAC42678.1"
/db_xref="GI:14536714"
/translation="MESISMGPXSLSTCLPNGINGIKDKARKVTGVIGSGDFAKS
LTIIRLRGYYHVIGSRPKFASFPFHVVDVTHEDALTNTNIFVAIHRHYISLW
DURHLVGLIDVSNMRINOYPESNAEYLASLPFDSLIVKGFNVVSAWALQIGPKD
ASRQYVICNNIQOQVIELARQNLNFIPIDIGSSSAREIENLPLRLFTLRGPPVV
AISLATFFLYSFVRDVIHPYARNOQSDYFKIPIEVNKTLPFIVAITLLSLVYLGLL
AAAYOLYGYTKYRRPMPWLETWLOCRKOLGLLSFFFAVHVAVSLCLPMRRERYLFL
NWAYOOVANIENSNEEVEWRIEMYISFGIMSLGLLSLLAVTSIPSPYNALNWRFEES
FQISLIGYVALLISFHVLLYGWKRAFEERYFRTPPNFVNLVLPISVILDLQLC
RYPD

BASE COUNT 648 a 537 c 520 g 748 t
ORIGIN
Query Match 78.2%; Score 315; DB 6; Length 2453;
Best Local Similarity 100.0%; Pred. NO. 4.2e-160;
Matches 315; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccagattcttgatgtcaaggatttaattgttctcagcttgaggcattcagttagg 60
|||||
Db 765 CCAGATTCCTTCATGTCACAGGATTAATGTGTCTCAGCTGGGCACCTTCAGTTAGG 824
QY 61 actaagatgacagccgagcaggtttatatatgcagcaacaatatattcaagcgcaaca 120
|||||
Db 825 ACCTAAGAGTGCAGCGCAGGCTTTATATATGCAGCAACAATATCAAGCGCGCAACA 884
QY 121 ggtattgaactgcccagcttgattcattccattgacttgatccttatcatc 180
|||||
Db 885 GGTATTGAATTCGCCGCGAGTTGAATTCATTCCTCCATTGACCTGGGATCCTTATCATC 944
QY 181 agcagagagattgaaaattaccctacagactcttactctctgagagggccagtggt 240
|||||
Db 945 AGCCAGAGAGATTGAAAATTTACCCCTACGACTCTTTACTCTCGAGAGGGCCAGTGGT 1004
QY 241 ggtagctaaagcttgccacatttttttctcttattctcttctgtagagatgattacca 300
|||||
Db 1005 GGTAGCTAAGCTTGCCGACATTTTTTTCCTTTTATCTCTTGTGCAGAGATGTATTCA 1064
QY 301 tccatatgctagaaa 315
|||||
Db 1065 TCCATATGTAGTAAAA 1079

RESULT 2
AC021898 87401 bp DNA ~ linear HTG 13-JUL-2000
LOCUS Homo sapiens chromosome 15 clone RP11-407J8 map 15, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
AC021898
VERSION AC021898.2 GI:9119282
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 87401)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 15, clone RP11-407J8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 87401)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P.,
Dearellano,K., Dekar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelano,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,

McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6730901.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2445
Center clone name: 407_J_8

* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 901: contig of 901 bp in length
* 902 1001: gap of 100 bp
* 1002 1907: contig of 906 bp in length
* 1908 2007: gap of 100 bp
* 2008 2912: contig of 905 bp in length
* 2913 3012: gap of 100 bp
* 3013 3872: contig of 860 bp in length
* 3873 3972: gap of 100 bp
* 3973 4868: contig of 896 bp in length
* 4869 4968: gap of 100 bp
* 4969 5874: contig of 906 bp in length
* 5875 5974: gap of 100 bp
* 5975 6897: contig of 923 bp in length
* 6898 6997: gap of 100 bp
* 6998 7903: contig of 906 bp in length
* 7904 8003: gap of 100 bp
* 8004 8913: contig of 910 bp in length
* 8914 9013: gap of 100 bp
* 9014 9906: contig of 893 bp in length
* 9907 10006: gap of 100 bp
* 10007 10906: contig of 900 bp in length
* 10907 11006: gap of 100 bp
* 11007 11910: contig of 904 bp in length
* 11911 12010: gap of 100 bp
* 12011 12924: contig of 914 bp in length
* 12925 13024: gap of 100 bp
* 13025 13930: contig of 906 bp in length
* 13931 14030: gap of 100 bp
* 14031 14919: contig of 889 bp in length
* 14920 15019: gap of 100 bp
* 15020 15900: contig of 881 bp in length
* 15901 16000: gap of 100 bp
* 16001 16938: contig of 938 bp in length
* 16939 17038: gap of 100 bp
* 17039 17953: contig of 915 bp in length
* 17954 18053: gap of 100 bp
* 18054 18908: contig of 855 bp in length
* 18909 19008: gap of 100 bp
* 19009 19906: contig of 898 bp in length

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:52:20 : Search time 4356.39 Seconds
(without alignments)
1935.868 Million cell updates/sec

Title: US-09-802-520-5
Perfect score: 403

Sequence: 1 cccagattcttgattgtca.....ctttgtccccctagtatac 403

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pi.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	315	78.2	2453	6	AX155249
2	236	58.6	87401	2	AC021898
3	236	58.6	156214	9	HSAC002064
4	219	54.3	444	6	AX106433
5	219	54.3	444	6	AX140724
6	219	54.3	444	6	AX200584
7	219	54.3	444	6	AX267240
8	203	50.4	162928	2	AC104475
9	157	39.0	172915	2	AC099742
10	153	38.0	328	6	AX106431
11	153	38.0	328	6	AX140722
12	153	38.0	328	6	AX200582
13	153	38.0	328	6	AX267238
14	50	12.4	322	6	AX155253
15	26	6.5	205085	2	AC026813
16	26	6.5	227144	2	AC092404
17	25	6.2	170102	9	AC009237
18	22	5.5	163386	9	AC015723
19	22	5.5	169030	2	AC026549
20	22	5.5	180161	2	AC063948
21	22	5.5	181259	9	AC063951
22	22	5.5	185322	2	AC021053
23	22	5.5	185576	2	AC025691
24	22	5.5	186546	2	AC097064
25	22	5.5	239704	9	AC099791
26	21	5.2	44662	8	AP004477
27	20	5.0	26539	3	U64836
28	20	5.0	57220	2	AC100359
29	20	5.0	57779	2	AC103799
30	20	5.0	70957	2	AP004214
31	20	5.0	71016	2	AC015459
32	20	5.0	117349	9	HST1060K6
33	20	5.0	135095	2	AC084714
34	20	5.0	138264	2	AP004749
35	20	5.0	152560	9	AC011339
36	20	5.0	156900	2	AC092808
37	20	5.0	156922	2	AC021687
38	20	5.0	164604	2	AC027383
39	20	5.0	168911	2	AC016853
40	20	5.0	171145	9	AL161450
41	20	5.0	176693	2	AP001586
42	20	5.0	178362	9	AL355615
43	20	5.0	180905	2	AC011688
44	19	4.7	6029	8	AF191096
45	19	4.7	67542	9	AC026422

ALIGNMENTS

RESULT 1
AX155249 LOCUS
DEFINITION Sequence 7 from Patent WO0140276.
ACCESSION AX155249
VERSION AX155249.1 GI:14536713
KEYWORDS human.
SOURCE Homo sapiens

REFERENCE 1 (bases 1 to 2453)
AUTHORS Afar,D.E., Hubert,R.S., Raitano,A.B., Saffran,D.C., Mitchell,S.C., Faris,M., and Jakobovits,A.

TITLE Serpentine transmembrane antigens expressed in human prostate

JOURNAL Patent: WO 0140276-A 7 07-JUN-2001;

FEATURES
source location/Qualifiers
1..2453 /organism="Homo sapiens"
355..1719 /db_xref="taxon:9606"

Urogenesys, Inc. (US)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

linear PAT 22-JUN-2001

2453 bp DNA

Sequence 7 from Patent WO0140276.

AX155249

GI:14536713

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

linear PAT 22-JUN-2001

THIS PAGE BLANK (USPTO)


```

Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 276 taaaccagtaccagaatccaa 297
|||||
Db 37 TAAACAGTACCAGATCCAA 16

RESULT 14
BF440211
LOCUS
DEFINITION
  BS29000195303 Lewin Cattle Spleen cDNA clone
  BS29000195303 3', mRNA sequence.
ACCESSION
  BF440211
VERSION
  BF440211.1 GI:11500094
KEYWORDS
  EST.
SOURCE
  cow.
ORGANISM
  Bos taurus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
  Bovidae; Bovinae; Bos.
REFERENCE
  1 (bases 1 to 295)
  Band,M.R., Larson,J.H., Rebeiz,M., Green,C.A., Heyen,D.W., Donovan
  ,J., Windish,R., Steining,C., Mahyuddin,P., Womack,J. and Lewin
  ,H.A.
  An ordered comparative map of the cattle and human genomes
  Genome Res. 10 (9), 1359-1368 (2000)
  20442530
  Contact: Lewin, H. A.
  W. M. Keck Center for Comparative and Functional Genomics
  University of Illinois at Urbana-Champaign
  340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
  61801, USA
  Tel: 217 333 5998
  Fax: 217 244 5617
  Email: h-lewin@uiuc.edu
  Funding for Cattle EST sequencing was provided by the USDA National
  Research Initiative, Project No. 98-35205-6644, and a grant from
  the Japanese Ministry of Agriculture Fisheries and Forestry to
  H.A.Lewin and J.E.Womack. Base-calling/Quality scores: PHRED orm
  Washington University Genome Center.Vector-trimming: Cross_Match
  from Washington University Genome Center PHRAP suite. This sequence
  is vector free and at least 150bp in length.
  Insert Length: 295 Std Error: 0.00
  Seq primer: TAGACTCACTAGGCGGAAT
  High quality sequence stop: 295.
  Location/Qualifiers
    1..295
      /organism="Bos taurus"
      /strain="Angus"
      /db_xref="taxon:9913"
      /clone="BS29000195303"
      /clone_lib="Lewin Cattle Spleen"
      /sex="Female"
      /dev_stage="Adult"
      /note="Organ: Spleen; Vector: pBluescript SK(+/-)"
BASE COUNT      89 a      38 c      84 g      84 t
ORIGIN

Query Match      4.1%; Score 20; DB 10; Length 295;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 agttgctctgaattttt 121
|||||
Db 168 AGTTGCTCTGAATTTT 187

RESULT 15
AZ353102
LOCUS
DEFINITION
  1M0091B19R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
  AZ353102
  315 bp DNA linear GSS 29-SEP-2000
  1M0091B19R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
  AZ353102

clone UUGC1M0091B19 R, DNA sequence.
AZ353102
VERSION
  AZ353102.1 GI:10432339
KEYWORDS
  GSS.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 315)
  Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
  Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
  ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
  and Wright,D.,Weiss,R.
  Mouse whole genome scaffolding with paired end reads from 10kb
  plasmid inserts
  Unpublished (2000)
  Contact: Robert B. Weiss
  University of Utah Genome Center
  University of Utah
  Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
  84112, USA
  Tel: 801 585 5606
  Fax: 801 585 7177
  Email: ddunn@genetics.utah.edu
  Insert Length: 10000 Std Error: 0.00
  Plate: 0091 row: B column: 19
  Seq primer: CACACAGGAACAGCTATGACC
  Class: plasmid ends
  High quality sequence stop: 315.
  Location/Qualifiers
    1..315
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="UUGC1M0091B19"
      /clone_lib="Mouse 10kb plasmid UUGC1M library"
      /sex="Male"
      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
      /note="Vector: PWD42nv; Purified genomic DNA from M.
      musculus C57BL/6J (male) was obtained from the Jackson
      Laboratory Mouse DNA Resource
      (http://www.jax.org/resources/documents/dnares/). The DNA
      was hydrodynamically sheared by repeated passage through a
      0.005 inch orifice at constant velocity. The sheared DNA
      was blunt end-repaired with T4 DNA polymerase and T4
      polynucleotide kinase. Adaptor oligonucleotides were
      ligated to the blunt ends in high molar excess. The
      adaptor DNA was purified and size-selected for a 9.5 to
      10.5 kb range using preparative agarose gel
      electrophoresis. Vector DNA was prepared from a derivative
      of pWD42 (gil4732114|gb|AF129072.1), a copy-number
      inducible derivative of plasmid R1. The vector was ligated
      with adaptors complementary to the insert adaptors and
      purified. The sheared, adaptor mouse DNA was annealed to
      adaptor vector DNA, and transformed into
      chemically-competent E. coli XL10-Gold (Stratagene) cells
      and selected for ampicillin resistance."
BASE COUNT      55 a      72 c      43 g      145 t
ORIGIN

Query Match      4.1%; Score 20; DB 12; Length 315;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 313 ttcattattcccgattctt 332
|||||
Db 53 TTCATTATTCCGATTCTT 72

Search completed: September 20, 2002, 08:33:44
Job time: 8708 sec

```

```

AA957846/c
LOCUS          AA957846          255 bp      mRNA      linear      EST 04-JUL-1999
DEFINITION    UI-R-E1-fv-a-01-0-UI.s2 UI-R-E1 Rattus norvegicus cDNA clone
               UI-R-E1-fv-a-01-0-UI 3', mRNA sequence.
ACCESSION     AA957846
VERSION       AA957846.1 GI:4277736
KEYWORDS      EST.
SOURCE        Norway rat.
ORGANISM      Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
               Rattus.
REFERENCE     1 (bases 1 to 255)
AUTHORS      Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene
               discovery
JOURNAL       Genome Res. 6 (9), 791-806 (1996)
MEDLINE       9704477
COMMENT      On May 7, 1998 this sequence version replaced gi:3121541.
               Contact: Soares, MB
               Program for Rat Gene Discovery and Mapping
               University of Iowa
               451 Eckstein Medical Research Building Iowa City, IA 52242, USA
               Tel: 319 335 8250
               Fax: 319 335 9565
               Email: msoares@blue.weeg.uiowa.edu
               The sequence tag present in the cDNA between the NotI site and the
               oligo-dT track served to identify it as a clone from the normalized
               adult 8-Day-Embryo library. cDNA Library Preparation: M. Fatima
               Bonaldo, Ph.D. Clone distribution: clones will be available through
               Research Genetics This clone is also available through the
               I.M.A.G.E. Consortium at LUNL (info@image.llnl.gov). IMAGE
               ID=1780077
               Seq primer: M13 Forward
POLYA-No.     1.
               Location/Qualifiers
               1..255
               /organism="Rattus norvegicus"
               /strain="Sprague-Dawley"
               /db_xref="taxon:10116"
               /clone="UI-R-E1-fv-a-01-0-UI"
               /clone_lib="UI-R-E1"
               /dev_stages="adult"
               /lab_host="DH10B (Life Technologies)"
               /note="Vector: pT73b-Pac (Pharmacia) with a modified
               polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-E1
               library is a subtracted library derived from the UI-R-E0
               library. The UI-R-E0 library consisted of a mixture of
               individually tagged normalized libraries constructed from
               8, 12 and 18-day embryo. The tag is a string of 3-5
               nucleotides present between the Not I site and the
               oligo-dT track which allows identification of the library
               of origin of a clone within the mixture. The subtracted
               library (UI-R-E1) was constructed as follows: PCR
               amplified cDNA inserts from a pool of UI-R-E0 clones from
               which 3' ESTs had been derived was used as a driver in a
               hybridization with the UI-R-E0 library in the form of
               single-stranded circles. The remaining single-stranded
               circles (subtracted library) was purified by
               hydroxyapatite column chromatography, converted to
               double-stranded circles and electroporated into DH10B
               bacteria (Life Technologies) to generate the UI-R-E1
               library. This procedure has been previously described
               (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
               1996)"
BASE COUNT    58 a 54 c 56 g 87 t
ORIGIN
Query Match   4.5%; Score 22; DB 9; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 taaaccagtaccagaatccaa 297
Db 37 TAAACCAAGTACCAGAATCCAA 16

RESULT 13
AA957889/c
LOCUS          AA957889          343 bp      mRNA      linear      EST 09-MAR-1999
DEFINITION    UI-R-E1-fv-a-01-0-UI.s1 UI-R-E1 Rattus norvegicus cDNA clone
               UI-R-E1-fv-a-01-0-UI 3', mRNA sequence.
ACCESSION     AA957889
VERSION       AA957889.1 GI:4277779
KEYWORDS      EST.
SOURCE        Norway rat.
ORGANISM      Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
               Rattus.
REFERENCE     1 (bases 1 to 343)
AUTHORS      Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE        Normalization and subtraction: two approaches to facilitate gene
               discovery
JOURNAL       Genome Res. 6 (9), 791-806 (1996)
MEDLINE       9704477
COMMENT      On May 7, 1998 this sequence version replaced gi:3121584.
               Contact: Soares, MB
               Program for Rat Gene Discovery and Mapping
               University of Iowa
               451 Eckstein Medical Research Building Iowa City, IA 52242, USA
               Tel: 319 335 8250
               Fax: 319 335 9565
               Email: msoares@blue.weeg.uiowa.edu
               The sequence tag present in the cDNA between the NotI site and the
               oligo-dT track served to identify it as a clone from the normalized
               adult 8-Day-Embryo library. cDNA Library Preparation: M. Fatima
               Bonaldo, Ph.D. Clone distribution: clones will be available through
               Research Genetics
               Seq primer: M13 Forward.
               Location/Qualifiers
               1..343
               /organism="Rattus norvegicus"
               /strain="Sprague-Dawley"
               /db_xref="taxon:10116"
               /clone="UI-R-E1-fv-a-01-0-UI"
               /clone_lib="UI-R-E1"
               /dev_stages="adult"
               /lab_host="DH10B (Life Technologies)"
               /note="Vector: pT73b-Pac (Pharmacia) with a modified
               polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-E1
               library is a subtracted library derived from the UI-R-E0
               library. The UI-R-E0 library consisted of a mixture of
               individually tagged normalized libraries constructed from
               8, 12 and 18-day embryo. The tag is a string of 3-5
               nucleotides present between the Not I site and the
               oligo-dT track which allows identification of the library
               of origin of a clone within the mixture. The subtracted
               library (UI-R-E1) was constructed as follows: PCR
               amplified cDNA inserts from a pool of UI-R-E0 clones from
               which 3' ESTs had been derived was used as a driver in a
               hybridization with the UI-R-E0 library in the form of
               single-stranded circles. The remaining single-stranded
               circles (subtracted library) was purified by
               hydroxyapatite column chromatography, converted to
               double-stranded circles and electroporated into DH10B
               bacteria (Life Technologies) to generate the UI-R-E1
               library. This procedure has been previously described
               (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
               1996)"
BASE COUNT    77 a 83 c 72 g 111 t
ORIGIN
Query Match   4.5%; Score 22; DB 9; Length 343;

```

ORIGIN

Query Match 4.7%; Score 23; DB 10; Length 515;
 Best Local Similarity 100.0%; Pred. No. 0.34;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 284 taccagaatccaatgctgaata 306
 |||||
 Db 428 TACCCAGATCCAATGCTGAATA 450

RESULT 10

LOCUS BG017798 516 bp mRNA linear EST 30-JAN-2001
 DEFINITION dabl0g01.y1 NICHD XGC Sp1 Xenopus laevis cDNA clone IMAGE:4174248
 5', mRNA sequence.

ACCESSION BG017798

VERSION BG017798.1 GI:12472384

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM

Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 Xenopodinae; Xenopus.

REFERENCE

1 (bases 1 to 516)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LLNL at: infoimage.llnl.gov

Seq primer: -40RP from Gibco

High quality sequence stop: 443.

FEATURES

source

1..516

Location/Qualifiers

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="IMAGE:4174248"

/clone_lib="NICHD XGC Sp1"

/dev_stage="adult"

/lab_host="DH10B (phage-resistant)"

/note="Organ: spleen; Vector: pCMV-SPT6; Site.1: NotI;

Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 2.4 kb. Constructed by Life

Technologies."

BASE COUNT 151 a 112 c 112 g 139 t 2 others

ORIGIN

Query Match 4.7%; Score 23; DB 10; Length 516;
 Best Local Similarity 100.0%; Pred. No. 0.34;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 284 taccagaatccaatgctgaata 306
 |||||
 Db 415 TACCCAGATCCAATGCTGAATA 437

RESULT 11

LOCUS AA997082 242 bp mRNA linear EST 09-MAR-1999
 DEFINITION UI-R-CO-hr-b-02-0-UI.s1 UI-R-CO Rattus norvegicus cDNA clone
 UI-R-CO-hr-b-02-0-UI 3', mRNA sequence.

ACCESSION AA997082

VERSION AA997082.1 GI:4281345

KEYWORDS EST.

SOURCE

ORGANISM

Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE 1 (bases 1 to 242)

AUTHORS

Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL

MEDLINE

COMMENT

Genome Res. 6 (9), 791-806 (1996)

97044477

On Jun 5, 1998 this sequence version replaced gi:3187943.

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the

oligo-dT track served to identify it as a clone from the normalized

adult 8-Day-Embryo library. cDNA Library Preparation: M. Fatima

Bonaldo, Ph.D. Clone distribution: clones will be available through

Research Genetics

Seq primer: M13 Forward.

Location/Qualifiers

1..242

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-CO-hr-b-02-0-UI"

/clone_lib="UI-R-CO"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-CO

library is a subtracted library derived from the UI-R-Al

and UI-R-El libraries. The UI-R-Al library consisted of a

mixture of individually tagged normalized libraries

constructed from rat placenta, adult lung, brain, liver,

kidney, heart, spleen, ovary, and muscle. The UI-R-El

library consisted of a mixture of individually tagged

normalized libraries constructed from 8, 12 and 18-day

embryo. The tag is a string of 3-5 nucleotides present

between the Not I site and the oligo-dT track which

allows identification of the library of origin of a clone

within the mixture. The subtracted library (UI-R-CO) was

constructed as follows: PCR amplified cDNA inserts from a

pool of UI-R-Al and UI-R-El clones from which 3' ESTs had

been derived was used as a driver in a hybridization with

the pooled UI-R-Al and UI-R-El library in the form of

single-stranded circles. The remaining single-stranded

circles (subtracted library) was purified by

hydroxyapatite column chromatography, converted to

double-stranded circles and electroporated into DH10B

bacteria (Life Technologies) to generate the UI-R-CO

library. This procedure has been previously described

(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,

1996)"

BASE COUNT 55 a 49 c 53 g 85 t

ORIGIN

Query Match 4.5%; Score 22; DB 9; Length 242;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 276 taaccagatcaccagaatccaa 297
 |||||
 Db 37 TAAACCAAGTACCAGAAATCCAA 16

RESULT 12

FEATURES

```

source
1. .561
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-brx-g-01-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=pineal-glands
TAG_SEQ=CAGAC"

```

```

BASE COUNT 154 a 115 c 129 g 163 t
ORIGIN

```

```

Query Match 5.3%; Score 26; DB 10; Length 561;
Best Local Similarity 100.0%; Pred. No. 0.0086;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 401 caggtttatatgcagcaacaatat 426
|||||
Db 250 CAGGTTTATATATGCAGCAACAATAT 225

```

RESULT 8

```

BG019037
LOCUS
DEFINITION
dab03c08.y1 NICHHD XGC Sp1 Xenopus laevis cDNA clone IMAGE:4173303
5', mRNA sequence.
ACCESSION
BG019037
VERSION
BG019037.1 GI:12474842
KEYWORDS
EST.
SOURCE
Xenopus laevis
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 502)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabps-f@mail.nih.gov
Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.

```

```

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LNL at: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 449.

```

FEATURES

```

source
1. .502
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:4173303"
/clone_lib="NICHHD XGC Sp1"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: spleen; Vector: pCMV-SPORT6; Site_1: Not I;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.4 kb. Constructed by Life
Technologies."

```

```

BASE COUNT 153 a 108 c 104 g 136 t 1 others
ORIGIN

```

```

Query Match 4.7%; Score 23; DB 10; Length 502;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 284 taccagaatccaatgctgaata 306
|||||
Db 439 TACCAGAATCCAATGCTGAATA 461

```

RESULT 9

```

BG657069
LOCUS
DEFINITION
dab27g09.y2 NICHHD XGC Sp1 Xenopus laevis cDNA clone IMAGE:4175801
5', mRNA sequence.
ACCESSION
BG657069
VERSION
BG657069.1 GI:13796858
KEYWORDS
EST.
SOURCE
African clawed frog.
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 515)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Other ESTs: dab27g09.x1
Contact: Robert Strausberg, Ph.D.
Email: cgabps-f@mail.nih.gov
Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LNL at: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 442.

```

FEATURES

```

source
1. .515
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:4175801"
/clone_lib="NICHHD XGC Sp1"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: spleen; Vector: pCMV-SPORT6; Site_1: Not I;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.4 kb. Constructed by Life
Technologies."

```

```

BASE COUNT 157 a 113 c 108 g 137 t

```

/clone_lib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (g14732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

Query Match	5.7%	Score 28;	DB 12;	Length 712;
Best Local Similarity	100.0%;	Pred. No.	0.00075;	
Matches	28;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	213	tgtgggaacctgagacatctgcttgtggg	240	
Dd	217	TGTGGGACCTGAGACATCTGCTTGTGGG	190	

RESULT	7
BI133492/C	
LOCUS	BI133492 561 bp mRNA linear EST 02-JUL-2001
DEFINITION	UI-M-BH3-brx-g-01-0-UI.sl NIH-BMAP_M_54 Mus musculus cDNA clone UI-M-BH3-brx-g-01-0-UI 3', mRNA sequence.

BI133492	561 bp	mRNA	linear	EST 02-JUL-2001
LOCUS				
DEFINITION	UI-M-BH3-brx-g-01-0-UI.s1 NIH_BMAP_M_54 Mus musculus cDNA clone			
	UI-M-BH3-brx-g-01-0-UI 3', mRNA sequence.			
ACCESSION	BI133492			
VERSION	BI133492.1			
KEYWORDS	GI:14583740			
SOURCE	EST.			
ORGANISM	house mouse.			
	Mus musculus.			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 561)			
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.			
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery			
JOURNAL	Genome Res. 6 (9), 791-806 (1996)			
MEDLINE	97044477			

Email: mestr@marl.mim.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NoRT site and the oligo-dT track served to identify it as a clone from the pineal glands tissue cDNA Library preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements

Seq primer: M13 Forward
POLYA=Yes.

Query Match	5.1%	Score 28;	DB 9;	Length 541;
Best Local Similarity	100.0%;	pred. No. 0.00072;		
Matches 28;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"

```

```
/clone="D930007L06"  
/clone_lib="RIKEN full-length enriched, 15 days embryo
```


/db_xref="taxon:9606"
/clone_lib="E0211"
/dev_stage="Adult"
/note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 144 a 100 c 99 g 140 t
ORIGIN

Query Match 79.5%; Score 392; DB 10; Length 483;
Best Local Similarity 100.0%; Pred. No. 6.5e-199;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 59 attatgctgcctatcgtgctataggagtagaatacctaaagtttgccttctgaattt 118
|||||
DB 483 ATTAGATCGGCTATCATGTGTCATAGGAAGTAGAAATCCTTAAGTTTCTTCTGAATTT 424
QY 119 ttctctcatgtgtagatgctactcatgatgaatgctctcacaaaaacaaataaata 178
|||||
DB 423 TTTCCTCATGTGAGTGTCACTCATCATGAAGATGCTCTCACAAAACAAATATAATA 364
QY 179 ttgttgctatcacagagaacattataccctccctgtggacctgagacatctgctgtg 238
|||||
DB 363 TTTGTTGCTATACACAGAGAACATATATACCTCCCTGTGGACCTGAGACATCTGCTTGTG 304
QY 239 ggtaaatcctcatgtatgagcgaataacatgagagataaaaccagtagaccagaatccaat 298
DB 303 GGTAAATCCTGATTGATGTGAGCAATAACATGAGGATAAACCAGTACCAGCAATCCAAT 244
QY 299 gctgaattttgcttcatttccacagattcttggattgtcaaggatttaagtgtgc 358
DB 243 GCTGAATATTGGCTTCATTATTCACAGATCTTTGATGTCAAGGATTTAATGTGTCT 184
QY 359 tcagcttggtgaccttcagttaggaacctaaaggatgcagccgagcaggttttatatgcagc 418
DB 183 TCAGCTTGGGCACCTTCAGTTAGGACCTTAAGGATGCCAGCGCAGGTTTATATATATATGCAGC 124
QY 419 acaatatattcaagcgcgaacacaggttattga 450
|||||
DB 123 AACAAATATTCAAGCGCGCAACACAGGTTATTGA 92

RESULT 2
A0888667
LOCUS nw77b11.s1 NCI_CGAP_Pr12 Homo sapiens cDNA clone IMAGE:1252605, EST 31-MAR-1998
DEFINITION mRNA sequence.
ACCESSION A0888667
VERSION A0888667.1 GI:3004342
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 180)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D., Rodrigo F. Chuquiqui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40ml3 fwd. Et from Amersham.

www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40ml3 fwd. Et from Amersham.
Location/Qualifiers
1..180
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1252605"
/clone_lib="NCI_CGAP_Pr12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from metastatic prostate lesion of the bone, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Krizman, NIH."
BASE COUNT 56 a 31 c 37 g 56 t
ORIGIN

Query Match 10.5%; Score 52; DB 9; Length 180;
Best Local Similarity 100.0%; Pred. No. 8.5e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 401 caggtttatatgcagcaacaattatccaagcgcgaacacaggtttatgaac 452
|||||
DB 129 CAGGTTTATATGTCAGCAACAATATTTCAAGCGCGCAACAGGTTATTGAAC 180

RESULT 3
A0935972
LOCUS ny30h06.s1 NCI_CGAP_Pr12 Homo sapiens cDNA clone IMAGE:1273307, EST 28-APR-1998
DEFINITION mRNA sequence.
ACCESSION A0935972
VERSION A0935972.1 GI:3093129
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 181)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D., Rodrigo F. Chuquiqui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40ml3 fwd. Et from Amersham.

Location/Qualifiers
1..181
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1273307"
/clone_lib="NCI_CGAP_Pr12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from metastatic prostate lesion of the bone, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Krizman, NIH."
BASE COUNT 56 a 34 c 35 g 56 t
ORIGIN

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 08:33:40 ; Search time 5053.42 Seconds
(without alignments)
1316.732 Million cell updates/sec

Title: US-09-802-520-4

Perfect score: 493

Sequence: 1 ggtcacttagtggtgattg.....ttcccatgacttggtatcc 493

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	392	79.5	483	10	BF883142 OV3-ET021
2	52	10.5	180	9	AA888667 nw77b11.s
3	52	10.5	181	9	AA935972 ny30h06.s
4	28	5.7	541	9	AI747886 ul03ell.y
5	28	5.7	680	9	BB603988 BB603988
c 6	28	5.7	712	12	AZ428664 IM0212F14
c 7	26	5.3	561	10	BI133492 UI-M-BH3-
8	23	4.7	502	10	BG019037 dab03c08
9	23	4.7	515	10	BG57069 dab27g09
10	23	4.7	516	10	BG017798 dab10g01
c 11	22	4.5	242	9	AA937082 UI-R-C0-h
c 12	22	4.5	255	9	AA957846 UI-R-El-f
c 13	22	4.5	343	9	AA957889 UI-R-El-f
14	20	4.1	295	10	BF440211 BS2900019
15	20	4.1	315	12	AZ353102 LM0091B19
16	20	4.1	336	10	BF560333 UI-R-El-f
17	20	4.1	370	12	AZ930320 474.dh#52

18	20	4.1	372	12	AZ238931	AZ238931	RPCI-23-7
19	20	4.1	379	9	AI930312	AI930312	ul63b08.y
c 20	20	4.1	495	9	AV893436	AV893436	AV893436
21	20	4.1	506	9	AI930004	AI930004	ul60d05.y
22	20	4.1	512	12	AQ912082	AQ912082	nbeb0016A
23	20	4.1	547	12	AQ591257	AQ591257	HS_2122_B
24	20	4.1	644	12	BH537545	BH537545	BQGR61TF
25	20	4.1	677	10	BI105662	BI105662	602892429
26	20	4.1	721	12	AZ019656	AZ019656	RPCI-23-3
27	20	4.1	760	12	AZ045243	AZ045243	nbeb0005IM
28	20	4.1	941	12	AG090249	AG090249	Pan trogl
29	19	3.9	109	12	AQ248591	AQ248591	T5K8-T7 T
30	19	3.9	250	9	AV264258	AV264258	AV264258
31	19	3.9	285	10	BM494429	BM494429	IPCGBR1_1
c 32	19	3.9	295	12	AQ245273	AQ245273	HS_2057_B
c 33	19	3.9	298	9	AI169414	AI169414	EST15264
c 34	19	3.9	349	9	AA381169	AA381169	EST94301
c 35	19	3.9	408	9	AV703667	AV703667	AV703667
c 36	19	3.9	421	10	BF944929	BF944929	CM1-NN021
c 37	19	3.9	425	10	BI292748	BI292748	UI-R-DO0-
c 38	19	3.9	426	10	BF282857	BF282857	EST447448
c 39	19	3.9	430	9	AA121811	AA121811	zn95ell.r
40	19	3.9	431	10	T95226	T95226	ye44f06.r1
41	19	3.9	462	10	BI473954	BI473954	fp49c10.y
c 42	19	3.9	479	10	BI292339	BI292339	UI-R-DO0-
43	19	3.9	486	10	BI326162	BI326162	fp86f07.y
c 44	19	3.9	499	10	BG894896	BG894896	355672.MA
c 45	19	3.9	506	9	AI600200	AI600200	EST251903

ALIGNMENTS

RESULT 1

BF883142/c

LOCUS

DEFINITION

OV3-ET0211-071200-529-b10 ET0211 Homo sapiens

ACCESSION

BF883142

VERSION

BF883142.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 483)

AUTHORS

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE

20202663

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=OV3&t2=OV3-ET0211-071200-529-b10&t3=2000-12-07&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 26

High quality sequence stop: 483.

Location/Qualifiers

1..483

/organism="Homo sapiens"

THIS PAGE BLANK (USPTO)

APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/686,878A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-686-878A-47

Query Match 3.2%; Score 16; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 gcttggtgggtaaaatc 247
|||||
DB 112 GCTTGTGGGTAAATC 97

RESULT 13
US-08-721-489-1/c
Sequence 1, Application US/08721489
Patent No. 5786465
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavallie, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/721,489
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-721-489-1

Query Match 3.2%; Score 16; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 gcttggtgggtaaaatc 247
|||||
DB 112 GCTTGTGGGTAAATC 97

RESULT 14
US-09-004-838-68
Sequence 68, Application US/09004838
Patent No. 6350933
GENERAL INFORMATION:
APPLICANT: Michelmore, Richard W.
APPLICANT: Meyers, Blake
TITLE OF INVENTION: Procedures and Materials for
TITLE OF INVENTION: Conferring Pest Resistance in Plants
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/004,838
FILING DATE: 09-JAN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/781,734
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-078810US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/991,942
;; FILING DATE: 09-SEP-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/304,309
;; FILING DATE: 09-SEP-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Smith, Timothy L.
;; REGISTRATION NUMBER: 35,367
;; REFERENCE/DOCKET NUMBER: 15280-210
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 543-9600
;; TELEFAX: (415) 543-5043
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4447 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 88..3162
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 1..4447
;; OTHER INFORMATION: /product= "Pig DPD"
US-08-991-942-3

Query Match 3.7%; Score 18; DB 3; Length 4447;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 164 aaacacaaataataattt 181
Db 3258 AAACAAATATATATT 3275

RESULT 8
US-08-687-080-115
; Sequence 115, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880

;; TELEFAX: (415) 324-0960
;; INFORMATION FOR SEQ ID NO: 115:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 7705 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; INDIVIDUAL ISOLATE: 3' END OF RAD50 GENOMIC SEQUENCE
US-08-687-080-115

Query Match 3.4%; Score 17; DB 2; Length 7705;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 135 atgtcactcatcga 151
Db 7390 ATGTCACATCATGAA 7406

RESULT 9
US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"

; CURRENT APPLICATION NUMBER: US/09/439,313
 ; CURRENT FILING DATE: 1999-11-12
 ; NUMBER OF SEQ ID NOS: 575
 ; SOFTWARE: FASTSEQ for Windows Version 3.0
 ; SEQ ID NO 212
 ; LENGTH: 328
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(328)
 ; OTHER INFORMATION: n = A,T,C or G
 ; US-09-439-313-212

Query Match 28.2%; Score 139; DB 4; Length 328;
 Best Local Similarity 100.0%; Pred. No. 3e-66;
 Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 327 attcttgattgcaaggatttaattgtctcagcttggcgacttcagttaggacctta 386
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 44 attcttgattgcaaggatttaattgtctcagcttggcgacttcagttaggacctta 103
 QY 387 aggatgcagccgagcaggtttatatatgcagcaacaattattcaagcgcgacaacaggtta 446
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 104 aggatgcagccgagcaggtttatatatgcagcaacaattattcaagcgcgacaacaggtta 163
 QY 447 ttgaacttgccgcaggtt 465
 ||||||||||||||||||||
 Db 164 ttgaacttgccgcaggtt 182

RESULT 5
 US-09-165-241-2/c
 ; Sequence 2, Application US/09165241
 ; Patent No. 6130077
 ; GENERAL INFORMATION:
 ; APPLICANT: Henry Yue
 ; APPLICANT: Karl J. Guegler
 ; APPLICANT: Neil C. Corley
 ; APPLICANT: Janice Au-Young
 ; TITLE OF INVENTION: HUMAN CYTOCHROME P450
 ; FILE REFERENCE: PF-0608 US
 ; CURRENT APPLICATION NUMBER: US/09/165,241
 ; CURRENT FILING DATE: 1998-10-01
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 2
 ; LENGTH: 1648
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE: -
 ; OTHER INFORMATION: 991729
 ; US-09-165-241-2

Query Match 3.9%; Score 19; DB 3; Length 1648;
 Best Local Similarity 100.0%; Pred. No. 0.66;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 114 aatttttcctcatgtggt 132
 ||||||||||||||||
 Db 413 AATTTTTCCTCATGTGTT 395

RESULT 6
 US-08-304-309-3
 ; Sequence 3, Application US/08304309
 ; Patent No. 5856454
 ; GENERAL INFORMATION:
 ; APPLICANT: GONZALEZ, Frank J.
 ; APPLICANT: FERNANDEZ-SALGUERO, Pedro
 ; TITLE OF INVENTION: CLONING AND EXPRESSION OF CDNA FOR HUMAN

; TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE
 ; NUMBER OF SEQUENCES: 13
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend Kourie and Crew
 ; STREET: Stuart Street Tower, One Market Plaza
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 94105-1493
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/304,309
 ; FILING DATE: 09-SEP-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Smith, Timothy L.
 ; REGISTRATION NUMBER: 35,367
 ; REFERENCE/DOCKET NUMBER: 15280-210
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 543-9600
 ; TELEFAX: (415) 543-5043
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4447 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 88..3162
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 1..4447
 ; OTHER INFORMATION: /product= "Pig DPD"
 ; US-08-304-309-3

Query Match 3.7%; Score 18; DB 2; Length 4447;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 164 aaacaaataataatttt 181
 ||||||||||||||||
 Db 3258 AAAACAAATATAATATT 3275

RESULT 7
 US-08-991-942-3
 ; Sequence 3, Application US/08991942
 ; Patent No. 6015673
 ; GENERAL INFORMATION:
 ; APPLICANT: GONZALEZ, Frank J.
 ; APPLICANT: FERNANDEZ-SALGUERO, Pedro
 ; TITLE OF INVENTION: CLONING AND EXPRESSION OF CDNA FOR HUMAN
 ; TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE
 ; NUMBER OF SEQUENCES: 13
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend Kourie and Crew
 ; STREET: Stuart Street Tower, One Market Plaza
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 94105-1493
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25

QY 345 gatttaattgtctcagcttgggcaacttcagtttaggacctaagatgcccagccgagcagg 404
Db 61 GATTTAAATTTCTCAGCTTGGCACTTCACTTAGGACCTAAGGATGCCAGCGGCAGG 120
QY 405 ttatatatgcagcaacaatatcagcgcaacaagcttattgaacttgcgcgcagc 464
Db 121 TTTATATATGCGACACAAATATTCAGCGCGACACAGGTTATTGACCTTGCCTGCCCACT 180

QY 465 tgaatttcattccattgacttgggatcc 493
Db 181 TGAATTTTCATTCCTCATTTGACTTGGGATCC 209

RESULT 2

US-09-439-313-214
; Sequence 214, Application US/09439313

; Patent No. 6329505

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan Louise

; APPLICANT: Jiang Yuqui

; APPLICANT: Reed, Steven G.

; APPLICANT: Kalos, Michael

; APPLICANT: Fanger, Gary

; APPLICANT: Retter, Mark

; APPLICANT: Solk, John

; APPLICANT: Day, Craig

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

; FILE REFERENCE: 210121.427C9

; CURRENT APPLICATION NUMBER: US/09/439,313

; CURRENT FILING DATE: 1999-11-12

; NUMBER OF SEQ ID NOS: 575

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 214

; LENGTH: 444

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(444)

; OTHER INFORMATION: n = A,T,C or G

US-09-439-313-214

Query Match 42.4%; Score 209; DB 4; Length 444;

Best Local Similarity 100.0%; Pred. No. 2.4e-104; Mismatches 0; Indels 0; Gaps 0;

Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 acccagaatccaatgctgaatttgggttcatttccagatttccagattctttgattgtcaag 344

Db 1 acccagaatccaatgctgaatttgggttcatttccagatttccagattctttgattgtcaag 60

QY 345 gatttaattgtctcagcttgggcaacttcagtttaggacctaagatgcccagccgagcagg 404

Db 61 gatttaattgtctcagcttgggcaacttcagtttaggacctaagatgcccagccgagcagg 120

QY 405 ttatatatgcagcaacaatatcagcgcaacaagcttattgaacttgcgcgcagc 464

Db 121 ttatatatgcagcaacaatatcagcgcaacaagcttattgaacttgcgcgcagc 180

QY 465 tgaatttcattccattgacttgggatcc 493

Db 181 tgaatttcattccattgacttgggatcc 209

RESULT 3

US-09-030-607-212

; Sequence 212, Application US/09030607

; Patent No. 6262245

; GENERAL INFORMATION:

;

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 212:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-09-030-607-212

Query Match 28.2%; Score 139; DB 4; Length 328;

Best Local Similarity 100.0%; Pred. No. 3e-66;

Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 attctttgattgcaagatttaattgtctcagcttgggcaacttcagtttaggaccta 386

Db 44 ATTCTTTGATTGCAAGGATTTAATGTCTCAGCTGGGCACCTCAGTTAGGACCTA 103

QY 387 aggatgccagcgccaggtttatatatgcagcaacaatatcagcgcaacaagctta 446

Db 104 AGGATGCCAGCGCGCAGGTTTATATATGCAGCAACAATATTCAGCGCGCAACAGGTTA 163

QY 447 ttgaacttgcgcgcagtt 465

Db 164 TTGAAGTTGCCCGCCAGTT 182

RESULT 4

US-09-439-313-212

; Sequence 212, Application US/09439313

; Patent No. 6329505

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan Louise

; APPLICANT: Jiang Yuqui

; APPLICANT: Reed, Steven G.

; APPLICANT: Kalos, Michael

; APPLICANT: Fanger, Gary

; APPLICANT: Retter, Mark

; APPLICANT: Solk, John

; APPLICANT: Day, Craig

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

; FILE REFERENCE: 210121.427C9

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:49:23 ; Search time 139.75 Seconds
(without alignments)
866.528 Million cell updates/sec

Title: US-09-802-520-4
Perfect score: 493
Sequence: 1 ggtcactgagtggtgattg.....ttccattgacttgggatcc 493

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 383533 seqs, 122816752 residues

Word size : 0
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCFUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	209	42.4	444	4	US-09-030-607-214
2	209	42.4	444	4	US-09-439-313-214
3	139	28.2	328	4	US-09-030-607-212
4	139	28.2	328	4	US-09-439-313-212
5	19	3.9	1648	3	US-09-165-241-2
6	18	3.7	4447	2	US-08-304-309-3
7	18	3.7	4447	3	US-08-991-942-3
8	17	3.4	7705	2	US-08-687-080-115
9	17	3.4	246240	2	US-08-724-394A-20
10	17	3.4	246240	2	US-08-724-394A-21
11	17	3.4	246240	2	US-08-724-394A-22
12	16	3.2	374	1	US-08-686-878A-47
13	16	3.2	374	1	US-08-721-489-1
14	16	3.2	461	4	US-09-004-838-68
15	16	3.2	792	1	US-08-403-545-6
16	16	3.2	792	4	US-08-404-381-6
17	16	3.2	1240	1	US-08-869-506-1
18	16	3.2	1240	3	US-09-128-967-1
19	16	3.2	2340	4	US-09-450-852-3
20	16	3.2	2540	2	US-08-511-485-3
21	16	3.2	2540	3	US-09-392-580-1
22	16	3.2	2946	4	US-09-175-928-3
23	16	3.2	3420	1	US-08-117-491-25
24	16	3.2	3420	1	US-08-271-364A-6
25	16	3.2	3420	2	US-08-222-715B-25
26	16	3.2	3796	1	US-08-343-760A-1
27	16	3.2	4707	1	US-08-004-139B-2

c 28	16	3.2	4707	2	US-08-811-492-2	Sequence 2, Appli
c 29	16	3.2	4707	5	PCT-US96-10545A-2	Sequence 2, Appli
c 30	16	3.2	5232	3	US-09-212-971-3	Sequence 3, Appli
c 31	16	3.2	5232	3	US-08-800-929A-3	Sequence 3, Appli
c 32	16	3.2	5232	4	US-09-617-053A-3	Sequence 3, Appli
c 33	16	3.2	6027	2	US-08-968-542C-1	Sequence 1, Appli
34	16	3.2	6122	1	US-08-403-545-1	Sequence 1, Appli
35	16	3.2	6122	4	US-08-404-381-1	Sequence 1, Appli
36	16	3.2	6328	4	US-08-913-832A-1	Sequence 1, Appli
37	16	3.2	152331	3	US-09-128-155-16	Sequence 16, Appli
c 38	16	3.2	176373	3	US-09-128-155-17	Sequence 17, Appli
c 39	15	3.0	383	5	PCT-US91-06039-1	Sequence 1, Appli
40	15	3.0	566	4	US-09-328-111-455	Sequence 455, App
c 41	15	3.0	852	2	US-09-069-330-1	Sequence 1, Appli
42	15	3.0	894	2	US-08-467-963C-28	Sequence 28, Appli
43	15	3.0	894	2	US-08-838-189D-28	Sequence 28, Appli
44	15	3.0	894	3	US-08-852-344D-28	Sequence 28, Appli
45	15	3.0	894	3	US-08-344-639E-28	Sequence 28, Appli

ALIGNMENTS

RESULT 1
US-09-030-607-214
; Sequence 214, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 214:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 444 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-030-607-214

Query Match 42.4%; Score 209; DB 4; Length 444;
Best Local Similarity 100.0%; Pred. No. 2.4e-104;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 285 accagaatccaatgctgaatttggcttcattattccagattctttgattgccaag 344
Db 1 ACCCAGATCCAAATGCTGAATTTGGCTTCATTATCCAGATTCCTTTGATTGCAAG 60

THIS PAGE BLANK (USPTO)

Query Match 82.2%; Score 405; DB 22; Length 539;
Best Local Similarity 100.0%; Pred. No. 1.4e-203;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggtcactgtagggtgattggaagtggagattttgccaataatccttgaccattgcacttat 60
Db 123 ggtcactgtagggtgattggaagtggagattttgccaataatccttgaccattgcacttat 182

Qy 61 tagatcggtctatcatgtgtcataggaagttagaatacctaaagtttgccttgaatttt 120
Db 183 tagatcggtctatcatgtgtcataggaagttagaatacctaaagtttgccttgaatttt 242

Qy 121 tcctcatgtgtgagatgcactcatcatgaagtgcctcacaataatataattatt 180
Db 243 tcctcatgtgtgagatgcactcatcatgaagtgcctcacaataatataattatt 302

Qy 181 tggctgtatcacagagaacattataacctccctgtggacctgagacatctgcttgggg 240
Db 303 tggctgtatcacagagaacattataacctccctgtggacctgagacatctgcttgggg 362

Qy 241 taaatccttgattgattgagcaataacatgaggataaaacagaccaggaatccaatgc 300
Db 363 taaatccttgattgattgagcaataacatgaggataaaacagaccaggaatccaatgc 422

Qy 301 tgaatatttggcttcatttccagatttcttgcagatttcttgcagatttcttgcctc 360
Db 423 tgaatatttggcttcatttccagatttcttgcagatttcttgcagatttcttgcctc 482

Qy 361 agcttgggaccttcagttaggaacctaaagatgccagccggcaggt 405
Db 483 agcttgggaccttcagttaggaacctaaagatgccagccggcaggt 527

RESULT 15
AAI49113
ID AAI49113 standard; DNA: 539 BP.
XX
AC AAI49113;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #17799 used to measure gene expression in human placenta sample.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
DR Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX

PS

xx Claim 25; SEQ ID NO 17799; 654pp; English.

CC

CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.

XX

SQ Sequence 539 BP; 156 A; 96 C; 118 G; 169 T; 0 other;

Query Match

82.2%; Score 405; DB 22; Length 539;

Best Local Similarity

100.0%; Pred. No. 1.4e-203;

Matches 405; Conservative

0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggtcactgtagggtgattggaagtggagattttgccaataatccttgaccattgcacttat 60

Db 123 ggtcactgtagggtgattggaagtggagattttgccaataatccttgaccattgcacttat 182

Qy 61 tagatcggtctatcatgtgtcataggaagttagaatacctaaagtttgccttgaatttt 120

Db 183 tagatcggtctatcatgtgtcataggaagttagaatacctaaagtttgccttgaatttt 242

Qy 121 tcctcatgtgtgagatgcactcatcatgaagtgcctcacaataatataattatt 180

Db 243 tcctcatgtgtgagatgcactcatcatgaagtgcctcacaataatataattatt 302

Qy 181 tggctgtatcacagagaacattataacctccctgtggacctgagacatctgcttgggg 240

Db 303 tggctgtatcacagagaacattataacctccctgtggacctgagacatctgcttgggg 362

Qy 241 taaatccttgattgattgagcaataacatgaggataaaacagaccaggaatccaatgc 300

Db 363 taaatccttgattgattgagcaataacatgaggataaaacagaccaggaatccaatgc 422

Qy 301 tgaatatttggcttcatttccagatttcttgcagatttcttgcagatttcttgcctc 360

Db 423 tgaatatttggcttcatttccagatttcttgcagatttcttgcagatttcttgcctc 482

Qy 361 agcttgggaccttcagttaggaacctaaagatgccagccggcaggt 405

Db 483 agcttgggaccttcagttaggaacctaaagatgccagccggcaggt 527

Search completed: September 20, 2002, 09:59:47

Job time: 11705 sec

```
Db 363 taaatccctgattgtagcaataacatgaggataaacaccagtagcccaagaatccaatgc 422
Qy 301 tgaatattgggttcattatccagattcttggattggtcaagagatttaattgtgtctc 360
Db 423 tgaatattgggttcattatccagattcttggattggtcaagagatttaattgtgtctc 482
Qy 361 agcttgggcacttcagttaggacctaagagatccagccgagcaggt 405
Db 483 agcttgggcacttcagttaggacctaagagatccagccgagcaggt 527

RESULT 13
AAK43038
ID AAK43038 standard; DNA; 539 BP.
XX
AC AAK43038;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 17595.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 17595; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 539 BP; 156 A; 96 C; 118 G; 169 T; 0 other;

Query Match 82.2%; Score 405; DB 22; Length 539;
Best Local Similarity 100.0%; Pred. No. 1.4e-203;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggtcactagtgtagtggaagtgagatttccaaatccttgaccattcgacttat 60
Db 123 ggtcactagtgtagtggaagtgagatttccaaatccttgaccattcgacttat 182
Qy 61 tagatgcggctatcatgtggtcataggaagtagaatactgaattgtcttgaatttt 120
Db 183 tagatgcggctatcatgtggtcataggaagtagaatactgaattgtcttgaatttt 242
```

```
Qy 121 tctcatgtggtagatgtcactcatcatgaagatgctctcacaaaaacaaataataatt 180
Db 243 tctcatgtggtagatgtcactcatcatgaagatgctctcacaaaaacaaataataatt 302
Qy 181 ttgtctatacacagagaacattatacctccctgtgtggacctgagacatctctgtgtgg 240
Db 303 ttgtctatacacagagaacattatacctccctgtgtggacctgagacatctctgtgtgg 362
Qy 241 taaaatccctgattgtgagcaataacatgaggataaaaccagtagcccaagaatccaatgc 300
Db 363 taaaatccctgattgtgagcaataacatgaggataaaaccagtagcccaagaatccaatgc 422
Qy 301 tgaatattgggttcattatccagattcttggattggtcaagagatttaattgtgtctc 360
Db 423 tgaatattgggttcattatccagattcttggattggtcaagagatttaattgtgtctc 482
Qy 361 agcttgggcacttcagttaggacctaagagatccagccgagcaggt 405
Db 483 agcttgggcacttcagttaggacctaagagatccagccgagcaggt 527

RESULT 14
AAI23801
ID AAI23801 standard; DNA; 539 BP.
XX
AC AAI23801;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #13734 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 13734; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 539 BP; 156 A; 96 C; 118 G; 169 T; 0 other;
```


KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ss.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 178..1650
FT /*tag= a
FT /product= "STMP1"
XX
XX WO200172962-A2.
XX
XX 04-OCT-2001.
XX
XX 23-MAR-2001; 2001WO-US09410.
XX
XX 24-MAR-2000; 2000US-191929P.
XX
XX (SAAT/) SAATCIOGLU F.
XX
XX Saatioglu F;
XX
XX WPI; 2001-662926/76.
XX P-PSDB; AAU10187.
XX
XX New polynucleotide for the diagnosis, prevention and treatment for
XX prostate and testis disorders, particularly prostate cancer, comprises
XX prostate-specific or testis-specific nucleic acids -
XX
XX Claim 5; Fig 4D; 114pp; English.
XX
XX The invention relates to substantially pure prostate-specific or
XX testis-specific polypeptides and the nucleic acids encoding them.
XX Also included are vectors and host cells expressing the proteins, a
XX transgenic animal expressing the protein, antibodies against the
XX proteins, probes for detecting the nucleic acids, antisense molecules
XX for the nucleic acids and methods of isolating modulators of the
XX proteins. Compounds that modulate the prostate specific or testis
XX specific polypeptide are useful to diagnose, prevent or treat disorders
XX of the testis or prostate particularly prostate cancer, benign
XX of the testis or prostate particularly prostate cancer, benign
XX prostatic hyperplasia, acute prostatitis, testicular cancer,
XX cryptorchidism, undescended, retractile, ascending or vanished
XX testis. Other proliferative disorders for which the modulators may be
XX used include lymphoma, leukemia, melanoma, ovarian cancer, breast
XX cancer, pancreatic cancer, liver cancer and lung cancer. The
XX present sequence represents the open reading frame of a prostate
XX specific protein, Six-transmembrane protein of Prostate 1, STMP1.
XX
XX Sequence 4329 BP; 1315 A; 817 C; 790 G; 1407 T; 0 other;

Query Match 100.0%; Score 493; DB 22; Length 4329;
Best Local Similarity 100.0%; Pred. No. 5.4e-250;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggtcactgtagtgtgattggaagtggagatttgcgaataccttgaccattcgacttat 60
Db 267 ggtcactgtagtgtgattggaagtggagatttgcgaataccttgaccattcgacttat 326
Qy 61 tagatcgcgctatcatgtgtgctataggaagttagaataccttaagtgttgcctgaatttt 120
Db 327 tagatcgcgctatcatgtgtgctataggaagttagaataccttaagtgttgcctgaatttt 386
Qy 121 tccatcattggttagatgtcactcatcatgaagatgctctcacaaaaacataataatt 180
Db 387 tccatcattggttagatgtcactcatcatgaagatgctctcacaaaaacataataatt 446
Qy 181 tttgtctatacacagaaacattatacctcctgtgagactgagacatctgcttggg 240
Db 447 tttgtctatacacagaaacattatacctcctgtgagactgagacatctgcttggg 506
Qy 241 taaaatcctgattgtgagcaataacatgaggataaacacagtcaccagaatccaatgc 300

Db 507 taaaatcctgattgtgagcaataacatgaggataaacacagtcaccagaatccaatgc 566
Qy 301 tgaatatttgcttcattattccacagattcttggattgcaaaaggatttaattgtcttc 360
Db 567 tgaatatttgcttcattattccacagattcttggattgcaaaaggatttaattgtcttc 626
Qy 361 agcttgggcaacttcagtttagacctaagatgccagccgaggtttatatatgcagcaa 420
Db 627 agcttgggcaacttcagtttagacctaagatgccagccgaggtttatatatgcagcaa 686
Qy 421 caatattcaagcgcagacaacaggttattgaacttgcgcaggttggaatttcattcccat 480
Db 687 caatattcaagcgcagacaacaggttattgaacttgcgcaggttggaatttcattcccat 746
Qy 481 tgacttgggatcc 493
Db 747 tgacttgggatcc 759
RESULT 9
ABA50938
ID ABA50938 standard; DNA; 539 BP.
XX AC
XX ABA50938;
XX
XX 01-FEB-2002 (first entry)
XX Human breast cell single exon nucleic acid probe #9633.
XX Human; microarray; single exon probe; gene expression; breast;
XX disease; cancer; ss.
XX Homo sapiens.
XX WO200157271-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00662.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 4; SEQ ID NO 9633; 327pp + sequence listing; English.
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting
XX the probes with a collection of detectably labelled nucleic acids
XX derived from mRNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias

CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,
 CC cryptorchidism, undescended, retractile, ascending or vanished
 CC testis. Other proliferative disorders for which the modulators may be
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The
 CC present sequence represents the second open reading frame of a prostate
 CC specific protein, Six-Transmembrane Protein of Prostate 1, STWPL.
 XX
 SQ Sequence 2238 BP; 607 A; 457 C; 453 G; 721 T; 0 other;

Query Match 100.0%; Score 493; DB 22; Length 2238;
 Best Local Similarity 100.0%; Pred. No. 5.3e-250;
 Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggctactgtggtgattggaagtgagagattttgccaatccttgaccattcgacttat 60
 Db 277 ggctactgtggtgattggaagtgagagattttgccaatccttgaccattcgacttat 336
 Oy 61 tagatcggtctatcatgtgtcattagagaagtagaataccttaagtttcttctgaattttt 120
 Db 337 tagatcggtctatcatgtgtcattagagaagtagaataccttaagtttcttctgaattttt 396
 Oy 121 tctcatgtgttagatgtcactcatcatgaagatgctctcacaaaaacaaataataatt 180
 Db 397 tctcatgtgttagatgtcactcatcatgaagatgctctcacaaaaacaaataataatt 456
 Oy 181 tgtgtctatacacagagaacattatacctcctgtggtgacctgagacatctgtgtggg 240
 Db 457 tgtgtctatacacagagaacattatacctcctgtggtgacctgagacatctgtgtggg 516
 Oy 241 taaaatcctgatgtgagcaataacatgaggataaaccagtaaccagaaatccaatgc 300
 Db 517 taaaatcctgatgtgagcaataacatgaggataaaccagtaaccagaaatccaatgc 576
 Oy 301 tgaatatttggtctcattatccagattcttctgattgtccaaagatttaattgtgtctc 360
 Db 577 tgaatatttggtctcattatccagattcttctgattgtccaaagatttaattgtgtctc 636
 Oy 361 agcttgggaccttaagtaggaactaagtagtccagccgaggtttatatatgcagcaa 420
 Db 637 agcttgggaccttaagtaggaactaagtagtccagccgaggtttatatatgcagcaa 696
 Oy 421 caattaccagcgacacacaggttattgaacttgcccgccagttgaatttcattcccat 480
 Db 697 caattaccagcgacacacaggttattgaacttgcccgccagttgaatttcattcccat 756
 Oy 481 tgacttgggatcc 493
 Db 757 tgacttgggatcc 769

RESULT 6
 AAD07072
 ID AAD07072 standard; cDNA; 2453 BP.
 XX
 AC AAD07072;
 XX
 DT 06-AUG-2001 (first entry)

DE Human six transmembrane epithelial antigen of prostate-2 clone GTD3 cDNA.
 XX
 XX Human: cytostatic; antiproliferative; vaccine; gene therapy;
 KW six transmembrane epithelial antigen of the prostate-2; STEAP-2;
 KW chromosome 7q21; cancer; prostate; colon; bladder; lung; ovarian;
 KW pancreatic; ss.
 XX
 OS Homo sapiens.

XX Key Location/Qualifiers
 FH RBS 352..358
 FT

FT FT /*tag= a
 FT FT /note= "kozak region"
 FT FT 355..1719
 FT FT /*tag= b
 FT FT /product= "Human six transmembrane epithelial antigen
 FT FT of the prostate (STEAP)-2"
 FT FT 709..2073
 FT FT /*tag= c
 FT FT /product= "Human six transmembrane epithelial antigen
 FT FT of the prostate (STEAP)-2, alternative version"
 FT FT /note= "CDS does not include start and stop codon"
 FT FT /transl_except= (pos:1714..1722, aa:Asp-Ala)
 FT FT /transl_except= (pos:1834..1842, aa:Arg-Ser)
 FT FT /transl_except= (pos:1957..1965, aa:Glu-Gly)
 FT FT /transl_except= (pos:2050..2058, aa:Thr-Ser)
 FT FT /transl_except= (pos:2062..2070, aa:Asn-Phe)
 FT FT /note= "Inframe stop codon alters the reading frame"
 FT FT /partial

WO200140276-A2.

07-JUN-2001.

06-DEC-2000; 2000WO-US33040.

06-DEC-1999; 99US-0455486.

(UROC-) UROGENESYS INC.

XX Afar DEH, Hubert RS, Raitano AB, Safran DC, Mitchell SC, Faris M;
 PI Jakobovits A;

WPI: 2001-367804/38.

P-PSDB: AAE02781, AAE02841.

PT New STEAP (six transmembrane epithelial antigen of the prostate)
 PT proteins, expressed in human cancers, useful for detecting and treating
 PT cancer -

PS Claim 4; Fig 9A-9D; 187pp; English.

XX The present sequence is human six transmembrane epithelial antigen of
 CC the prostate (STEAP)-2 clone GTD3 cDNA. STEAP is a member of cell
 CC surface serpentine transmembrane antigens. STEAP-2 gene is located on
 CC chromosome 7q21 and is used in gene therapy. Inhibiting the development
 CC or progression of a cancer (eg. prostate, colon, bladder, lung, ovarian
 CC and pancreatic) expressing STEAP or inhibiting growth or killing cells
 CC expressing STEAP in a patient, comprises administering a vaccine
 CC composition to the patient. Treating a patient with a cancer that
 CC expresses STEAP, or inhibiting growth or killing cells expressing STEAP,
 CC comprises administering to the patient a vector encoding single chain
 CC monoclonal antibody that comprises the variable domains of the heavy and
 CC light chains of the monoclonal antibody that specifically binds to STEAP,
 CC such that the vector delivers the single chain monoclonal antibody coding
 CC sequence to the cancer cells and the encoded single chain monoclonal
 CC antibody is expressed intracellularly.
 CC Note: The present sequence is also shown in sequence listing of the
 CC specification, but it lacks nucleotides at its 5' end.

XX Sequence 2453 BP; 648 A; 537 C; 520 G; 748 T; 0 other;

Query Match 100.0%; Score 493; DB 22; Length 2453;
 Best Local Similarity 100.0%; Pred. No. 5.4e-250;
 Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ggctactgtggtgattggaagtgagagattttgccaatccttgaccattcgacttat 60
 Db 444 ggctactgtggtgattggaagtgagagattttgccaatccttgaccattcgacttat 503

Oy 61 tagatcggtctatcatgtgtcattagagaagtagaataccttaagtttcttctgaatttt 120
 Db 504 tagatcggtctatcatgtgtcattagagaagtagaataccttaagtttcttctgaatttt 563

XX Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.
KW
OS Homo sapiens.
XX WO200175067-A2.
PN 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX P-PSDB; ABG12306.
DR
XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
PT
XX Claim 1; SEQ ID No 12297; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2192 BP; 559 A; 507 C; 551 G; 575 T; 0 other;

Query Match 100.0%; Score 493; DB 23; Length 2192;
Best Local Similarity 100.0%; Pred. NO. 5.3e-250;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggtcactgtaggtgtgattggaagtggagattttgccaatactctgaccattcgactt 60
DB 384 ggtcactgtaggtgtgattggaagtggagattttgccaatactctgaccattcgactt 443
QY 61 tagatgcggtatcatgtgtgcataggaagttagaataccttaagtgtgcttgaatttt 120
DB 444 tagatgcggtatcatgtgtgcataggaagttagaataccttaagtgtgcttgaatttt 503
QY 121 tctcatgtgttagatgtctcaactcatgatgaagtgtctcacaacaaataataatt 180
DB 504 tctcatgtgttagatgtctcaactcatgatgaagtgtctcacaacaaataataatt 563
QY 181 ttttctatacacagagaacattatcctcctgtgtggaccctgagacatctgttggg 240
DB 564 ttttctatacacagagaacattatcctcctgtgtggaccctgagacatctgttggg 623

QY 241 taaaatcctgattgattgagcaataacatgagagataaacaccagatcaccagaatccaatgc 300
DB 624 taaaatcctgattgattgagcaataacatgagagataaacaccagatcaccagaatccaatgc 683
QY 301 tgaatatttggtcttatttccacagattctttgattgtcacaaggatttaattgtctc 360
DB 684 tgaatatttggtcttatttccacagattctttgattgtcacaaggatttaattgtctc 743
QY 361 agcttgggcacttcagttaggtacctaaggtgccagccgcaggtttatatatgcagcaa 420
DB 744 agcttgggcacttcagttaggtacctaaggtgccagccgcaggtttatatatgcagcaa 803
QY 421 caatattcaagcgcacacacaggttattgaacttgcgcaggttgaatttcattcccat 480
DB 804 caatattcaagcgcacacacaggttattgaacttgcgcaggttgaatttcattcccat 863
QY 481 tgacttgggatacc 493
DB 864 tgacttgggatacc 876
RESULT 5
AAS15810
ID AAS15810 standard; cDNA; 2238 BP.
XX
AC AAS15810;
XX
DT 16-JAN-2002 (first entry)
XX
DE Human ORF2 of Six-Transmembrane Protein of Prostate 1, STMP1.
XX
DE Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer; benign prostatic hyperplasia; acute prostatitis; testicular cancer; cryptorchidism; testicular disorder; proliferative disorder; lymphoma; leukoemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer; liver cancer; lung cancer; cytostatic; ss; ORF2.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 188..1552
FT /tag- a "STMP1, ORF2"
FT /product- "STMP1, ORF2"
XX WO200172962-A2.
XX
PD 04-OCT-2001.
XX
XX 23-MAR-2001; 2001WO-US09410.
PF
XX
XX 24-MAR-2000; 2000US-191929P.
PR
XX (SAAT/) SAATCIOGLU F.
XX
XX Saatioglu F;
PI
XX
XX WPI; 2001-662926/76.
DR P-PSDB; AAU10188.
XX
PT New polynucleotide for the diagnosis, prevention and treatment for prostate and testis disorders, particularly prostate cancer, comprises prostate-specific or testis-specific nucleic acids -
PT
XX
PS Claim 5; Fig 4G; 114pp; English.
XX
XX The invention relates to substantially pure prostate-specific or testis-specific polypeptides and the nucleic acids encoding them. Also included are vectors and host cells expressing the proteins, a transgenic animal expressing the protein, antibodies against the proteins, probes for detecting the nucleic acids, antisense molecules for the nucleic acids and methods of isolating modulators of the proteins. Compounds that modulate the prostate specific or testis

Db 327 tagatcggtcatcatggtgcataggaagtgaatactctaaagttgtgcttgaatttt 386
Qy 121 tctcatgtggttagatgtcactcatcatgaagatgctctcacaacacacaaataatt 180
Db 387 tctcatgtggttagatgtcactcatcatgaagatgctctcacaacacacaaataatt 446
Qy 181 tgtgtctacacagagaaacattatactccctgtggacctgagacatctgtgtggg 240
Db 447 tgtgtctacacagagaaacattatactccctgtggacctgagacatctgtgtggg 506
Qy 241 taaatctctgattgtagagcaataacatgaggagataaacacagtaaccacgaatccaatgc 300
Db 507 taaatctctgattgtagagcaataacatgaggagataaacacagtaaccacgaatccaatgc 566
Qy 301 tgaatttggcttcattatccagattcttggattggtcacaagatttaattgtgtctc 360
Db 567 tgaatttggcttcattatccagattcttggattggtcacaagatttaattgtgtctc 626
Qy 361 agcttgggcacttcagttaggaacctaaagatgccagccggcaggtttatatatgcagcaa 420
Db 627 agcttgggcacttcagttaggaacctaaagatgccagccggcaggtttatatatgcagcaa 686
Qy 421 caattatcaagcgcagacaacaggttatgtgaacttgcgcgcaggttgaatttccattcccat 480
Db 687 caattatcaagcgcagacaacaggttatgtgaacttgcgcgcaggttgaatttccattcccat 746
Qy 481 tgaacttgggatcc 493
Db 747 tgaacttgggatcc 759

RESULT 3
AAS15813
ID AAS15813 standard; cDNA; 2102 BP.
AC AAS15813;
XX 16-JAN-2002 (first entry)
XX Human cDNA encoding ORF3 of Six-Transmembrane Protein of Prostate 1.
XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ss; ORF3.
OS Homo sapiens.
XX Key Location/Qualifiers
XX CDS 171..1430
XX FT /*tag= a
XX FT /product= "STMP1, ORF3"
XX PN WO200172962-A2.
XX XX 04-OCT-2001.
XX PF 23-MAR-2001; 2001WO-US09410.
XX PR 24-MAR-2000; 2000US-191929P.
XX XX (SAAT/) SAATCIOGLU F.
XX Saatioglu F;
XX WPI; 2001-662926/76.
XX P-PSDB; AAU10189.
XX New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids

XX Claim 4; Fig 4K; 114pp; English.
PS The invention relates to substantially pure prostate-specific or
XX testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence encodes prostate specific protein, Six-Transmembrane
XX Protein of Prostate 1, STMP1, ORF3.
SQ Sequence 2102 BP; 595 A; 424 C; 422 G; 661 T; 0 other;
Query Match 100.0%; Score 493; DB 22; Length 2102;
Best Local Similarity 100.0%; Pred. No. 5.3e-250;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ggtcactgtaggtgattggaagtggagattttcccaaatccttgacattgacttat 60
Db 260 ggtcactgtaggtgattggaagtggagattttcccaaatccttgacattgacttat 319
Qy 61 tagatcggtctatcatgtggtcataggaagttagaataccttaagttgcttgaatttt 120
Db 320 tagatcggtctatcatgtggtcataggaagttagaataccttaagttgcttgaatttt 379
Qy 121 tctcatgtggttagatgtcactcatcatgaagatgctctcacaacacacaaataatt 180
Db 380 tctcatgtggttagatgtcactcatcatgaagatgctctcacaacacacaaataatt 439
Qy 181 tgtgtctacacagagaaacattatactccctgtggacctgagacatctgtgtggg 240
Db 440 tgtgtctacacagagaaacattatactccctgtggacctgagacatctgtgtggg 499
Qy 241 taaatctctgattgtagagcaataacatgaggagataaacacagtaaccacgaatccaatgc 300
Db 500 taaatctctgattgtagagcaataacatgaggagataaacacagtaaccacgaatccaatgc 559
Qy 301 tgaatttggcttcattatccagattcttggattggtcacaagatttaattgtgtctc 360
Db 560 tgaatttggcttcattatccagattcttggattggtcacaagatttaattgtgtctc 619
Qy 361 agcttgggcacttcagttaggaacctaaagatgccagccggcaggtttatatatgcagcaa 420
Db 620 agcttgggcacttcagttaggaacctaaagatgccagccggcaggtttatatatgcagcaa 679
Qy 421 caattatcaagcgcagacaacaggttatgtgaacttgcgcgcaggttgaatttccattcccat 480
Db 680 caattatcaagcgcagacaacaggttatgtgaacttgcgcgcaggttgaatttccattcccat 739
Qy 481 tgaacttgggatcc 493
Db 740 tgaacttgggatcc 752
RESULT 4
AAS76493
ID AAS76493 standard; cDNA; 2192 BP.
XX XX
AC AAS76493;
XX 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #12297.
DE

```
XX  Saatcioglu F;
PI  WPI: 2001-662926/76.
DR  P-PSDB; AAU10188.
XX
PT  New polynucleotide for the diagnosis, prevention and treatment for
PT  prostate and testis disorders, particularly prostate cancer, comprises
PT  prostate-specific or testis-specific nucleic acids
XX
XX  Claim 4; Fig 4H; 114pp; English.
XX
XX  The invention relates to substantially pure prostate-specific or
XX  testis-specific polypeptides and the nucleic acids encoding them.
XX  Also included are vectors and host cells expressing the proteins, a
XX  transgenic animal expressing the protein, antibodies against the
XX  proteins, probes for detecting the nucleic acids, antisense molecules
XX  for the nucleic acids and methods of isolating modulators of the
XX  proteins. Compounds that modulate the prostate specific or testis
XX  specific polypeptide are useful to diagnose, prevent or treat disorders
XX  of the testis or prostate particularly prostate cancer, benign
XX  prostatic hyperplasia, acute prostatitis, testicular cancer,
XX  cryptorchidism, undescended, retractile, ascending or vanished
XX  testis. Other proliferative disorders for which the modulators may be
XX  used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
XX  cancer, pancreatic cancer, liver cancer and lung cancer. The
XX  present sequence encodes prostate specific protein, Six-transmembrane
XX  Protein of Prostate 1, STMP1, ORF2.
XX
XX  Sequence 1561 BP; 420 A; 319 C; 346 G; 476 T; 0 other;

Query Match      100.0%; Score 493; DB 22; Length 1561;
Best Local Similarity 100.0%; Pred. No. 5.3e-250;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 ggtcactgtagggtgattggaagtggagattttgccaaatccttgaccattcgacttat 60
Db  277 ggtcactgtagggtgattggaagtggagattttgccaaatccttgaccattcgacttat 336
QY  61 tagatcggtctatcattgtggtcataggaagtagaataccttaagttgcttctgaatttt 120
Db  337 tagatcggtctatcattgtggtcataggaagtagaataccttaagttgcttctgaatttt 396
QY  121 tctctatggttagatgtcactcatcatcatgaagatgctctccacaaacaaataataatt 180
Db  397 tctctatggttagatgtcactcatcatgaagatgctctccacaaacaaataataatt 456
QY  181 tgttgcatacacagagaacattatccctcctgtggtgacctgagacatcgtctgtgg 240
Db  457 tgttgcatacacagagaacattatccctcctgtggtgacctgagacatcgtctgtgg 516
QY  241 taaatccctgattgattgagcaataacatgagataaacacagatacccgaaatccaatgc 300
Db  517 taaatccctgattgattgagcaataacatgagataaacacagatacccgaaatccaatgc 576
QY  301 tgaattattggttcattatccagattcctttgattgattgattgattgattgattgatt 360
Db  577 tgaattattggttcattatccagattcctttgattgattgattgattgattgattgatt 636
QY  361 agcttgggcacttcattgtagcctaagatgccagccgaggttttatatgacgaa 420
Db  637 agcttgggcacttcattgtagcctaagatgccagccgaggttttatatgacgaa 696
QY  421 caatatcaagcgcgacacaggttattgaacttgcctccaggttgaatttcattcccat 480
Db  697 caatatcaagcgcgacacaggttattgaacttgcctccaggttgaatttcattcccat 756
QY  481 tgacttgggatcc 493
Db  757 tgacttgggatcc 769
```

RESULT 2

```
AA15802
ID  AAS15802 standard; cDNA; 1680 BP.
XX
AC  AAS15802;
XX
DT  16-JAN-2002 (first entry)
XX
DE  Human cDNA encoding Six-Transmembrane Protein of Prostate 1, STMP1.
XX
KW  Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW  benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW  cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW  leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW  liver cancer; lung cancer; cytostatic; ss.
XX
OS  Homo sapiens.
XX
FH  Key Location/Qualifiers
CDS 178..1650
     /*tag= a
     /product= "STMP1"
XX
XX  WO200172962-A2.
XX  04-OCT-2001.
XX  23-MAR-2001; 2001WO-US09410.
XX  24-MAR-2000; 2000US-191929P.
XX  (SAAT/) SAATCIOGLU F.
XX  Saatcioglu F;
XX  WPI: 2001-662926/76.
XX  P-PSDB; AAU10187.
XX  New polynucleotide for the diagnosis, prevention and treatment for
XX  prostate and testis disorders, particularly prostate cancer, comprises
XX  prostate-specific or testis-specific nucleic acids
XX
XX  Claim 4; Fig 4E; 114pp; English.
XX
XX  The invention relates to substantially pure prostate-specific or
XX  testis-specific polypeptides and the nucleic acids encoding them.
XX  Also included are vectors and host cells expressing the proteins, a
XX  transgenic animal expressing the protein, antibodies against the
XX  proteins, probes for detecting the nucleic acids, antisense molecules
XX  for the nucleic acids and methods of isolating modulators of the
XX  proteins. Compounds that modulate the prostate specific or testis
XX  specific polypeptide are useful to diagnose, prevent or treat disorders
XX  of the testis or prostate particularly prostate cancer, benign
XX  prostatic hyperplasia, acute prostatitis, testicular cancer,
XX  cryptorchidism, undescended, retractile, ascending or vanished
XX  testis. Other proliferative disorders for which the modulators may be
XX  used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
XX  cancer, pancreatic cancer, liver cancer and lung cancer. The
XX  present sequence encodes a prostate specific protein, Six-Transmembrane
XX  Protein of Prostate 1, STMP1.
XX
XX  Sequence 1680 BP; 467 A; 334 C; 373 G; 506 T; 0 other;
```

```
Query Match      100.0%; Score 493; DB 22; Length 1680;
Best Local Similarity 100.0%; Pred. No. 5.3e-250;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 ggtcactgtagggtgattggaagtggagattttgccaaatccttgaccattcgacttat 60
Db  267 ggtcactgtagggtgattggaagtggagattttgccaaatccttgaccattcgacttat 326
QY  61 tagatcggtctatcattgtggtcataggaagtagaataccttaagttgcttctgaatttt 120
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:59:43 ; Search time 600.75 Seconds
(without alignments)
1408.970 Million cell updates/sec

Title: US-09-802-520-4
Perfect score: 493
Sequence: 1 ggcactgagtgatg.....ttccattgacttgatcc 493

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 1736436 seqs, 858457221 residues
Word size : 0

Total number of hits satisfying chosen parameters: 3472872
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	493	100.0	1561	22 AAS15811	Human cDNA encodin
2	493	100.0	1680	22 AAS15802	Human cDNA encodin
3	493	100.0	2102	22 AAS15813	Human cDNA encodin
4	493	100.0	2192	23 AAS76493	DNA encoding novel
5	493	100.0	2238	22 AAS15810	Human ORF2 of Six-
6	493	100.0	2433	22 AAD07072	Human six transmem
7	493	100.0	3900	23 AAS64300	DNA encoding novel
8	493	100.0	4329	22 AAS15801	Human ORF of Six-T
9	405	82.2	539	22 ABA50938	Human breast cell

10	405	82.2	539	22 ABA68917	Human foetal liver
11	405	82.2	539	22 ABA35870	Probe #14336 for g
12	405	82.2	539	22 AAK17249	Human brain expres
13	405	82.2	539	22 AAK43038	Human bone marrow
14	405	82.2	539	22 AAI23801	Probe #13734 for g
15	405	82.2	539	22 AAI49113	Probe #17799 used
16	405	82.2	539	22 AAI09410	Probe #9401 used t
17	405	82.2	1725	22 AAS15793	Human DNA for Six-
18	405	82.2	1953	22 ABA45820	Human breast cell
19	405	82.2	1953	22 ABA56335	Human foetal liver
20	405	82.2	1953	22 ABA25976	Probe #4442 for ge
21	405	82.2	1953	22 AAK04514	Human brain expres
22	405	82.2	1953	22 AAK30016	Human bone marrow
23	405	82.2	1953	22 AAI14606	Probe #4539 for ge
24	405	82.2	1953	22 AAI35978	Probe #4664 used t
25	405	82.2	1953	22 AAI04420	Probe #4411 used t
26	403	81.7	525	22 AAS15796	Human Six-Transmem
27	403	81.7	525	22 AAS15805	Human Six-Transmem
28	209	42.4	444	19 AAV61348	Forward DNA sequen
29	209	42.4	444	19 AAV58684	Prostate tumour sp
30	209	42.4	444	21 AAO66447	Human immunogenic
31	209	42.4	444	22 AAS63655	Human prostate CDN
32	209	42.4	444	22 AAS10206	Human prostate tum
33	209	42.4	444	22 AAH93563	Human prostate-spe
34	209	42.4	444	22 AAH84877	Human prostate-spe
35	209	42.4	444	22 AAO26268	Prostate tumour an
36	139	28.2	328	19 AAV61346	Forward DNA sequen
37	139	28.2	328	19 AAV58682	Prostate tumour sp
38	139	28.2	328	21 AAO66445	Human immunogenic
39	139	28.2	328	22 AAS63653	Human prostate CDN
40	139	28.2	328	22 AAS10204	Human prostate tum
41	139	28.2	328	22 AAH93561	Human prostate-spe
42	139	28.2	328	22 AAH84875	Human prostate-spe
43	139	28.2	328	22 AAO26266	Prostate tumour an
44	90	18.3	528	22 AAS15797	Human Six-Transmem
45	90	18.3	528	22 AAS15806	Human Six-Transmem

ALIGNMENTS

RESULT 1
AAS15811
ID AAS15811 standard; cDNA; 1561 BP.
XX AAS15811;
AC AAS15811;
XX 16-JAN-2002 (first entry)
XX Human cDNA encoding ORF2 of Six-Transmembrane Protein of Prostate 1.
XX Human: Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ss; ORF2.
XX Homo sapiens.
XX Key Location/Qualifiers
FT CDS 188..1552
FT /*tag= a
FT /product= "STMP1, ORF2"
XX
PN WO200172962-A2.
XX
XX
PD 04-OCT-2001.
XX
XX
PF 23-MAR-2001; 2001WO-US09410.
XX
XX
PR 24-MAR-2000; 2000US-191929P.
XX
XX (SAAT/) SAATCIOGLU F.

THIS PAGE BLANK (USPTO)

```

misc_feature 1..26708
              /note="assembly_fragment
              clone_end:SP6
              vector_side:left"
misc_feature 26809..109655
              /note="assembly_fragment
              clone_end:T7
              vector_side:left"
misc_feature 109756..227144
              /note="assembly_fragment"
BASE COUNT 68114 a 46378 c 46791 g 65638 t 223 others
ORIGIN

```

```

Query Match      5.7%; Score 28; DB 2; Length 227144;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 213 tgtggacctgagacatctgtgtgtgg 240
      |||||
Db 127889 TGTGGACCTGAGACATCTGCTTGTGGG 127862

```

Search completed: September 20, 2002, 09:52:20
 Job time: 12998 sec

* 29020 29933: contig of 914 bp in length
* 29934 30033: gap of 100 bp
* 30034 30930: contig of 897 bp in length
* 30931 31030: gap of 100 bp
* 31031 31924: contig of 894 bp in length
* 31925 32024: gap of 100 bp
* 32025 32902: contig of 878 bp in length
* 32903 33002: gap of 100 bp
* 33003 33890: contig of 888 bp in length
* 33891 33990: gap of 100 bp
* 33991 34900: contig of 910 bp in length
* 34901 35000: gap of 100 bp
* 35001 35895: contig of 895 bp in length
* 35896 35995: gap of 100 bp
* 35996 36840: contig of 845 bp in length
* 36841 36940: gap of 100 bp
* 36941 37822: contig of 882 bp in length
* 37823 37922: gap of 100 bp
* 37923 38845: contig of 923 bp in length
* 38846 38945: gap of 100 bp
* 38946 39823: contig of 878 bp in length
* 39824 39923: gap of 100 bp
* 39924 40792: contig of 869 bp in length
* 40793 40892: gap of 100 bp
* 40893 41811: contig of 919 bp in length
* 41812 41911: gap of 100 bp
* 41912 42806: contig of 895 bp in length
* 42807 42906: gap of 100 bp
* 42907 43784: contig of 878 bp in length
* 43785 43884: gap of 100 bp
* 43885 44777: contig of 893 bp in length
* 44778 44877: gap of 100 bp
* 44878 45768: contig of 891 bp in length
* 45769 45868: gap of 100 bp
* 45869 46761: contig of 893 bp in length
* 46762 46861: gap of 100 bp
* 46862 47713: contig of 852 bp in length
* 47714 47813: gap of 100 bp
* 47814 48729: contig of 916 bp in length
* 48730 48829: gap of 100 bp
* 48830 49779: contig of 930 bp in length
* 49780 49879: gap of 100 bp
* 49880 50794: contig of 915 bp in length
* 50795 50894: gap of 100 bp
* 50895 51797: contig of 903 bp in length
* 51798 51897: gap of 100 bp
* 51898 52786: contig of 889 bp in length
* 52787 52886: gap of 100 bp
* 52887 53779: contig of 893 bp in length
* 53780 53879: gap of 100 bp
* 53880 54782: contig of 903 bp in length
* 54783 54882: gap of 100 bp
* 54883 55769: contig of 887 bp in length
* 55770 55869: gap of 100 bp
* 55870 56736: contig of 867 bp in length
* 56737 56836: gap of 100 bp
* 56837 57715: contig of 879 bp in length
* 57716 57815: gap of 100 bp
* 57816 58712: contig of 897 bp in length
* 58713 58812: gap of 100 bp
* 58813 59729: contig of 917 bp in length
* 59730 59829: gap of 100 bp
* 59830 60638: contig of 809 bp in length
* 60639 60738: gap of 100 bp
* 60739 61654: contig of 916 bp in length
* 61655 61754: gap of 100 bp
* 61755 62680: contig of 926 bp in length
* 62681 62780: gap of 100 bp
* 62781 63664: contig of 884 bp in length
* 63665 63764: gap of 100 bp
* 63765 64670: contig of 906 bp in length
* 64671 64770: gap of 100 bp
* 64771 65666: contig of 896 bp in length

* 65667 65766: gap of 100 bp
* 65767 66636: contig of 870 bp in length
* 66637 66736: gap of 100 bp
* 66737 67604: contig of 868 bp in length
* 67605 67704: gap of 100 bp
* 67705 68576: contig of 872 bp in length
* 68577 68676: gap of 100 bp
* 68677 69601: contig of 925 bp in length
* 69602 69701: gap of 100 bp
* 69702 70618: contig of 917 bp in length

Query Match 18.9%; Score 93; DB 2; Length 87401;
Best Local Similarity 100.0%; Pred. No. 1.7e-43;
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 401 caggttatatatcagcgaacaattattcaagcgcgacacacaggttattgaattgcccgc 460
|||||
DB 46422 CAGGTTTATATGAGCAACAATATTCAAGCGCGACACAGGTTATTGAAGTTGCCCGC 46481
|||||
QY 461 cagttgaattcattccattgacttggtatcc 493
|||||
DB 46482 CAGTTGAATTTTCATCCCATTTGATGGATCC 46514
|||||

RESULT 14
AC026813
LOCUS AC026813 205085 bp DNA linear HTG 27-OCT-2001
DEFINITION Mus musculus chromosome 5 clone RP23-119M19 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 10 unordered pieces.
ACCESSION AC026813
VERSION AC026813.2 GI:16506398
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 205085)
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Mastrian,S.D.,
McGloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Shevchenko,Y.,
Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Zhang,J.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 205085)
Green,E.D.
Direct Submission
Submitted (24-MAR-2000) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
On Oct 27, 2001 this sequence version replaced gi:7321458.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc_mouse@nhgri.nih.gov
----- Project Information
Center project name: xh
Center clone name: 119M19
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 199516 bases at least Q40
Consensus quality: 200596 bases at least Q30
Consensus quality: 201000 bases at least Q20
Insert size: 197000; agarose-fp
Insert size: 192000; pulse-field-gel
Insert size: 204185; sum-of-contigs
Quality coverage: 9.36x in Q20 bases; agarose-fp

```
Matches 334; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 68 ggcctatcatgtggtcactagaagaatcctaaagttgtctctgaatttttctcctat 127
Db 83379 GGCATCATGTCGTATAGAGAGTAGAATCTTAAGTTGCTTCGAATTTTTCCTCAT 83438
QY 128 gtggtagatgtcactcatcatgaagatgctctcacaaacaaataataattgttgt 187
Db 83439 GTGGTAGATGTACCATCATGAAGATGCTCTCACAAAAACAAATATAATATTGTGTCT 83498
QY 188 atacacagagaacattatattaccctccctgtggagacctgagacatctctgtgggtaaac 247
Db 83499 ATACATGAGAGACATATATACCTCCCTGTGGACCTTAAGACATCTGCTTGTGGTAAAAATC 83558
QY 248 ctgattgatgtgagcgaataacatgaggataaaacccagatccacagaatccaaatgctgaat 307
Db 83559 CTGATTGATGTGAGCAATAAACATGAGGATAAACACAGTACCCAGATCCCAATGCTGAATAT 83618
QY 308 ttgggttcatttccacagatcttttgattgtcacaaggatttaattgttctcagctgg 367
Db 83619 TTGGCGTCAATATTCCAGATTCTTTGATTGTCAAGAGGATTTAATGTTCTCTCAGCTGG 83678
QY 368 gcacttcagttaggacctgaagatgcacgcgcaggt 405
Db 83679 GCACCTCAGTTAGGACCTAAGGATGCCACCGCAGGT 83716

RESULT 13
AC021898      87401 bp      DNA      linear      HTG 13-JUL-2000
LOCUS Homo sapiens chromosome 15 clone RP11-407J8 map 15, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
ACCESSION AC021898
VERSION AC021898.2 GI:9119282
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 87401)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 15, clone RP11-407J8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 87401)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepeil,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrelra,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisan,R., Pollara,V., Raymond,B., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., VO,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jul 13, 2000 this sequence version replaced gi:6730901.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
```

```
----- Project Information
Center project name: L2445
Center clone name: 407_J_8
-----
* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 901: contig of 901 bp in length
* 902 1001: gap of 100 bp
* 1002 1907: contig of 905 bp in length
* 1908 2007: gap of 100 bp
* 2008 2912: contig of 905 bp in length
* 2913 3012: gap of 100 bp
* 3013 3872: contig of 860 bp in length
* 3873 3972: gap of 100 bp
* 3973 4868: contig of 895 bp in length
* 4869 4968: gap of 100 bp
* 4969 5874: contig of 906 bp in length
* 5875 5974: gap of 100 bp
* 5975 6897: contig of 923 bp in length
* 6898 6997: gap of 100 bp
* 6998 7903: contig of 906 bp in length
* 7904 8003: gap of 100 bp
* 8004 8913: contig of 910 bp in length
* 8914 9013: gap of 100 bp
* 9014 9906: contig of 893 bp in length
* 9907 10006: gap of 100 bp
* 10007 10906: contig of 900 bp in length
* 10907 11006: gap of 100 bp
* 11007 11910: contig of 904 bp in length
* 11911 12010: gap of 100 bp
* 12011 12924: contig of 914 bp in length
* 12925 13024: gap of 100 bp
* 13025 13930: contig of 906 bp in length
* 13931 14030: gap of 100 bp
* 14031 14919: contig of 889 bp in length
* 14920 15019: gap of 100 bp
* 15020 15900: contig of 881 bp in length
* 15901 16000: gap of 100 bp
* 16001 16938: contig of 938 bp in length
* 16939 17038: gap of 100 bp
* 17039 17953: contig of 915 bp in length
* 17954 18053: gap of 100 bp
* 18054 18908: contig of 855 bp in length
* 18909 19008: gap of 100 bp
* 19009 19906: contig of 898 bp in length
* 19907 20006: gap of 100 bp
* 20007 20922: contig of 916 bp in length
* 20923 21022: gap of 100 bp
* 21023 21936: contig of 914 bp in length
* 21937 22036: gap of 100 bp
* 22037 22942: contig of 906 bp in length
* 22943 23042: gap of 100 bp
* 23043 23928: contig of 886 bp in length
* 23929 24028: gap of 100 bp
* 24029 24911: contig of 883 bp in length
* 24912 25011: gap of 100 bp
* 25012 25887: contig of 876 bp in length
* 25888 25987: gap of 100 bp
* 25988 26883: contig of 896 bp in length
* 26884 26983: gap of 100 bp
* 26984 27912: contig of 929 bp in length
* 27913 28012: gap of 100 bp
* 28013 28919: contig of 907 bp in length
* 28920 29019: gap of 100 bp
```

```

|||||
Db 44 ATCTTTGATTGTCAAAGGATTATATGTTGCTCAGCTTGGCGACTTCAGTTAGGACCTA 103
|||||
Oy 387 aggtgacgagcgaggtttatattatgcagcaacaattatcaagcgacgaacagggtta 446
|||||
Db 104 AGGATGCCAGCGGAGGTTTATATATGTCAGCAACATATTCAAGCGGACACAGGTTA 163
|||||
Oy 447 ttgaacttgccgcgcagtt 465
|||||
Db 164 TTGAACCTTGCCCGCCAGTT 182
|||||

RESULT 11
LOCUS AX267238 328 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 212 from Patent WO0173032.
ACCESSION AX267238
VERSION AX267238.1 GI:16516016
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Kalos,M.D., Fanger,G.R., Retter,M.W., Stolk,J.A., Day,C.H.,
Vedvick,T.S., Carter,D., Li,S.X., Wang,A., Skeiky,Y.A., Hepler,W.T.
and Henderson,R.A.
TITLE Compositions and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0173032-A 212 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source
1..328
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 84 a 73 c 61 g 106 t 4 others
ORIGIN
Query Match 28.28; Score 139; DB 6; Length 328;
Best Local Similarity 100.0%; Pred. No. 4e-71;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 327 attctttgattgcaagagatttaattgtctcagcttggtgacacttcagttaggacctta 386
|||||
Db 44 ATCTTTGATTGTCAAAGGATTATATGTTGCTCAGCTTGGCGACTTCAGTTAGGACCTA 103
|||||
Oy 387 aggtgacgagcgaggtttatattatgcagcaacaattatcaagcgacgaacagggtta 446
|||||
Db 104 AGGATGCCAGCGGAGGTTTATATATGTCAGCAACATATTCAAGCGGACACAGGTTA 163
|||||
Oy 447 ttgaacttgccgcgcagtt 465
|||||
Db 164 TTGAACCTTGCCCGCCAGTT 182
|||||

RESULT 12
AC099742
LOCUS AC099742 172915 bp DNA linear HTG 20-NOV-2001
DEFINITION Papio cynocephalus anubis clone RP41-167P22, WORKING DRAFT
SEQUENCE, 4 unordered pieces.
ACCESSION AC099742
VERSION AC099742.1 GI:17017546
KEYWORDS HTG; HTGS_PASEI; HTGS_DRAFT.
SOURCE olive baboon.
ORGANISM Papio cynocephalus anubis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Papio.
REFERENCE 1 (bases 1 to 172915)
AUTHORS Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,

```

```

Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Shevchenko,Y.,
Stantipop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
Turgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Zhang,L.-H. and Green,E.D.
TITLE NISC Comparative Sequencing Initiative
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 172915)
Green,E.D.
Direct Submission
Submitted (20-NOV-2001) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nhgri.nih.gov
----- Project Information
Center project name: ccv
Center clone name: 167P22
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171178 bases at least Q40
Consensus quality: 171910 bases at least Q30
Consensus quality: 172236 bases at least Q20
Insert size: 130000; agarose-fp
Quality coverage: 12.58x in Q20 bases; agarose-fp
Quality coverage: 9.47x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2438: contig of 2438 bp in length
* 2439 2538: gap of unknown length
* 2539 8133: contig of 5595 bp in length
* 8134 8233: gap of unknown length
* 8234 40378: contig of 32145 bp in length
* 40379 40478: gap of unknown length
* 40479 172915: contig of 132437 bp in length.
FEATURES
source
1..172915
/organism="Papio cynocephalus anubis"
/db_xref="taxon:9555"
/clone="RP41-167P22"
/clone_lib="RP41"
1..2438
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
2539..8133
/note="assembly_fragment"
8234..40378
/note="assembly_fragment"
clone_end:T7
vector_side:right"
40479..172915
/note="assembly_fragment"
BASE COUNT 52509 a 31733 c 32277 g 56096 t 300 others
ORIGIN

Query Match 27.28; Score 134; DB 2; Length 172915;
Best Local Similarity 98.8%; Pred. No. 6.9e-68;

```

BASE COUNT 120 a 100 c 78 g 143 t 3 others
ORIGIN /db_xref="taxon:9606"

Query Match 42.4%; Score 209; DB 6; Length 444;
Best Local Similarity 100.0%; Pred. No. 8.1e-113;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 285 accagaatccaatgctgaatttggtcttcatttccagatcttctgattgtcaag 344
|||||
Db 1 ACCAGATCCAATGCTGAATTTGGCTTCATTTATCCAGATCTTTGATTGTCAAG 60
|||||

Qy 345 gatttaattgtctcagcttggtggcacttcagttaggacctgaagatgccagccggcagg 404
|||||
Db 61 GATTTAATGTTGCTCAGCTTGGGCACCTTCAGTTAGGACCTAAGGATGCCAGCGCAGG 120
|||||

Qy 405 ttatatatgcagcaacaattatcaagcgacacacaggttattgaacttgcgcgcaggt 464
|||||
Db 121 TTTATATATGACGACCAACATATTCGAAGCGCGACACAGGTTATTGAACCTTGCCCGCCAGT 180
|||||

Qy 465 tgaatttcattccattgacttggtatcc 493
|||||
Db 181 TGAATTTCCATCCATTCGACTTGGGATCC 209
|||||

RESULT 8
LOCUS AX106431 328 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 212 from Patent WO0125272.
ACCESSION AX106431
VERSION AX106431.1 GI:13922110
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 328)
AUTHORS Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.
TITLE Compositions and methods for the therapy and diagnosis of prostate cancer
JOURNAL Patent: WO 0125272-A 212 12-APR-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
misc_feature 1..328
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 84 a 73 c 61 g 106 t 4 others
ORIGIN /note="n = A,T,C or G"

Query Match 28.2%; Score 139; DB 6; Length 328;
Best Local Similarity 100.0%; Pred. No. 4e-71;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 attcttggatgtcaagaggatttaattgtctcagcttggtggcacttcagttaggacctta 386
|||||
Db 44 ATTCCTTTGATTGTCAAGGATTAAATGTTGCTCAGCTTGGGCACCTTCAGTTAGGACCTTA 103
|||||

Qy 387 aggatgccagccgaggtttatatatgcagcaacaattatccaagcgacacaggtta 446
|||||
Db 104 AGGATGCCAGCGCGCAGGTTTATATATATGCAGCAACAATATTCAAGCGCGACACAGGTTA 163
|||||

Qy 447 ttgaacttgcgcgcaggtt 465
|||||
Db 164 TTGAACCTTGCCCGCCAGTT 182
|||||

RESULT 9
LOCUS AX140722 328 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 212 from Patent WO0151633.
ACCESSION AX200582
VERSION AX200582.1 GI:15390402
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 328)
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,
Stolk,J.A., Skeiky,Y.A., Wang,A. and Meagher,M.J.
TITLE Compositions and methods for the therapy and diagnosis of prostate cancer
JOURNAL Patent: WO 0151633-A 212 19-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
misc_feature 1..328
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 84 a 73 c 61 g 106 t 4 others
ORIGIN

Sequence 212 from Patent WO0134802.
ACCESSION AX140722
VERSION AX140722.1 GI:14280833
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 328)
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,
Skelky,Y.A. and Wang,A.
TITLE Compositions and methods for the therapy and diagnosis of prostate cancer
JOURNAL Patent: WO 0134802-A 212 17-MAY-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
misc_feature 1..328
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 84 a 73 c 61 g 106 t 4 others
ORIGIN

Query Match 28.2%; Score 139; DB 6; Length 328;
Best Local Similarity 100.0%; Pred. No. 4e-71;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 attcttggatgtcaagaggatttaattgtctcagcttggtggcacttcagttaggacctta 386
|||||
Db 44 ATTCCTTTGATTGTCAAGGATTAAATGTTGCTCAGCTTGGGCACCTTCAGTTAGGACCTTA 103
|||||

Qy 387 aggatgccagccgaggtttatatatgcagcaacaattatccaagcgacacaggtta 446
|||||
Db 104 AGGATGCCAGCGCGCAGGTTTATATATATGCAGCAACAATATTCAAGCGCGACACAGGTTA 163
|||||

Qy 447 ttgaacttgcgcgcaggtt 465
|||||
Db 164 TTGAACCTTGCCCGCCAGTT 182
|||||

RESULT 10
LOCUS AX200582 328 bp DNA linear PAT 29-AUG-2001
DEFINITION Sequence 212 from Patent WO0151633.
ACCESSION AX200582
VERSION AX200582.1 GI:15390402
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 328)
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,
Stolk,J.A., Skeiky,Y.A., Wang,A. and Meagher,M.J.
TITLE Compositions and methods for the therapy and diagnosis of prostate cancer
JOURNAL Patent: WO 0151633-A 212 19-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
misc_feature 1..328
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 84 a 73 c 61 g 106 t 4 others
ORIGIN

Query Match 28.2%; Score 139; DB 6; Length 328;
Best Local Similarity 100.0%; Pred. No. 4e-71;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 327 attcttggatgtcaagaggatttaattgtctcagcttggtggcacttcagttaggacctta 386
|||||
Db 44 ATTCCTTTGATTGTCAAGGATTAAATGTTGCTCAGCTTGGGCACCTTCAGTTAGGACCTTA 103
|||||

Qy 387 aggatgccagccgaggtttatatatgcagcaacaattatccaagcgacacaggtta 446
|||||
Db 104 AGGATGCCAGCGCGCAGGTTTATATATATGCAGCAACAATATTCAAGCGCGACACAGGTTA 163
|||||

Qy 447 ttgaacttgcgcgcaggtt 465
|||||
Db 164 TTGAACCTTGCCCGCCAGTT 182
|||||

```
misc_feature 1..444
/notes="n = A,T,C or G"
BASE COUNT 120 a 100 c 78 g 143 t 3 others
ORIGIN

Query Match 42.4%; Score 209; DB 6; Length 444;
Best Local Similarity 100.0%; Pred. No. 8.1e-113;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 accagaatccaatgctgaattattggtcttcattccagattccagattcttggatgcaag 344
Db 1 ACCAGAATCCAATGCTGAATATTGGCTTCATTATCCAGATCTTTGATTGTCAAAG 60

QY 345 gatttaattgtctcagcttggtcagcttcagttgagacctaagatgccagccgcagc 404
Db 61 GATTTAATGTTGCTCAGCTTGGCAGCTTCAGTTAGACCTTAGAGATGCCAGCCGGCAGG 120

QY 405 ttatatatgcagacaataattcaagcgcgacacaggttattgaaacttgcgcgcagc 464
Db 121 TTTATATATGCAGACAATAATTCAAGCGGCACACAGGTATTGAACCTTGCCTGCCAGT 180

QY 465 tgaatttcattccattgacttggatcc 493
Db 181 TGAATTTCAATCCCATTTGACTTGGATCC 209

RESULT 5
AX140724
LOCUS AX140724 444 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 214 from Patent WO0134802.
ACCESSION AX140724
VERSION AX140724.1 GI:14280835
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 444)
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,
Skeiky,Y.A., Wang,A.
COMPOSITIONS and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0134802-A 214 17-MAY-2001;
CORIXA CORPORATION (US)
FEATURES
source 1..444
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 120 a 100 c 78 g 143 t 3 others
ORIGIN

Query Match 42.4%; Score 209; DB 6; Length 444;
Best Local Similarity 100.0%; Pred. No. 8.1e-113;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 accagaatccaatgctgaattattggtcttcattccagattcttggatgcaag 344
Db 1 ACCAGAATCCAATGCTGAATATTGGCTTCATTATCCAGATCTTTGATTGTCAAAG 60

QY 345 gatttaattgtctcagcttggtcagcttcagttgagacctaagatgccagccgcagc 404
Db 61 GATTTAATGTTGCTCAGCTTGGCAGCTTCAGTTAGACCTTAGAGATGCCAGCCGGCAGG 120

QY 405 ttatatatgcagacaataattcaagcgcgacacaggttattgaaacttgcgcgcagc 464
Db 121 TTTATATATGCAGACAATAATTCAAGCGGCACACAGGTATTGAACCTTGCCTGCCAGT 180

QY 465 tgaatttcattccattgacttggatcc 493
Db 181 TGAATTTCAATCCCATTTGACTTGGATCC 209

RESULT 6
AX200584
LOCUS AX200584 444 bp DNA linear PAT 29-AUG-2001
DEFINITION Sequence 214 from Patent WO0151633.
ACCESSION AX200584
VERSION AX200584.1 GI:15390404
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 444)
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,
Stolk,J.A., Skeiky,Y.A., Wang,A. and Meagher,M.J.
COMPOSITIONS and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0151633-A 214 19-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source 1..444
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 120 a 100 c 78 g 143 t 3 others
ORIGIN

Query Match 42.4%; Score 209; DB 6; Length 444;
Best Local Similarity 100.0%; Pred. No. 8.1e-113;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 accagaatccaatgctgaattattggtcttcattccagattcttggatgcaag 344
Db 1 ACCAGAATCCAATGCTGAATATTGGCTTCATTATCCAGATCTTTGATTGTCAAAG 60

QY 345 gatttaattgtctcagcttggtcagcttcagttgagacctaagatgccagccgcagc 404
Db 61 GATTTAATGTTGCTCAGCTTGGCAGCTTCAGTTAGACCTTAGAGATGCCAGCCGGCAGG 120

QY 405 ttatatatgcagacaataattcaagcgcgacacaggttattgaaacttgcgcgcagc 464
Db 121 TTTATATATGCAGACAATAATTCAAGCGGCACACAGGTATTGAACCTTGCCTGCCAGT 180

QY 465 tgaatttcattccattgacttggatcc 493
Db 181 TGAATTTCAATCCCATTTGACTTGGATCC 209

RESULT 7
AX267240
LOCUS AX267240 444 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 214 from Patent WO0173032.
ACCESSION AX267240
VERSION AX267240.1 GI:16516018
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Kalos,M.D., Fanger,G.R., Retter,M.W., Stolk,J.A., Day,C.H.,
Vedvick,T.S., Carter,D., Li,S.X., Wang,A., Skeiky,Y.A., Hepler,W.T.
and Henderson,R.A.
COMPOSITIONS and methods for the therapy and diagnosis of prostate
cancer
JOURNAL Patent: WO 0173032-A 214 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source 1..444
/organism="Homo sapiens"
```

```
RESULT 3
AC104475/c
LOCUS
DEFINITION Pan troglodytes clone RP43-120K11, WORKING DRAFT SEQUENCE, 6
AC104475 162928 bp DNA linear HTG 12-DEC-2001
unordered pieces.
AC104475
AC104475.1 GI:17530717
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE chimpanzee.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 (bases 1 to 162928)
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Akter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-Q.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Masiello, C., Mastrian, S.D.,
McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Stantiripop, S.,
Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurgon, C., Vogt, J.L.,
Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and
Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 162928)
Green, E.D.
Direct Submission
Submitted (12-DEC-2001) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
Center: NIH Intramural Sequencing Center
Genome Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@hgri.nih.gov
----- Project Information
Center project name: ces
Center clone name: 120K11
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 161001 bases at least Q40
Consensus quality: 161422 bases at least Q30
Consensus quality: 161562 bases at least Q20
Insert size: 143000; agarose-fp
Insert size: 162428; sum-of-contigs
Quality coverage: 10.76x in Q20 bases; agarose-fp
Quality coverage: 9.47x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 13922: contig of 13922 bp in length
* 13923 14022: gap of unknown length
* 14023 36248: contig of 22226 bp in length
* 36249 36349: gap of unknown length
* 36349 66192: contig of 29844 bp in length
* 66193 66292: gap of unknown length
* 66293 92168: contig of 25876 bp in length
* 92169 92268: gap of unknown length
* 92269 122037: contig of 29768 bp in length
* 122037 122137: gap of unknown length
* 122137 162928: contig of 40792 bp in length.
* Location/Qualifiers
* 1. 162928
* /organism="Pan troglodytes"

/db_xref="taxon:9598"
/clone="RP43-120K11"
/clone_lib="RP43"
1. 13922
/note="assembly_fragment"
clone_end:17
vector_side:left"
14023. 36248
/note="assembly_fragment"
36349. 66192
/note="assembly_fragment"
66293. 92168
/note="assembly_fragment"
92269. 122036
/note="assembly_fragment"
122137. 162928
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
BASE COUNT 51528 a 29878 c 30533 g 50485 t 504 others
ORIGIN

Query Match 61.7%; Score 304; DB 2; Length 162928;
Best local Similarity 99.7%; Pred. No. 3.6e-169;
Matches 354; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ggtcactgtagtgtgagtggaagtgagatttgcgaatccttgaccattcgacttat 60
|||||
Db 37789 GGTCACTGTAGTGTGATTTGGAAGTGGAGATTTTGGCAATCTCTGACCATTCGACTTAT 37730
|||||

QY 61 tagatgcggtcatcatgtgtgcataggaagtagaataccttaagtgtgcttctgaatttt 120
|||||
Db 37729 TAGATGTGGCTATCATGTGGTCATAGGAAGTAGAAATCTAAGTTTGTCTTCTGAATTTT 37670
|||||

QY 121 tctcatgtgttagatgtcaatcatcatgaatgaatgagatgctctcacaaaaataataatt 180
|||||
Db 37669 TCCTCATGTGGTAGATGTGTCATCATGAGATGCTCTCACAAAAACAATAATAATAT 37610
|||||

QY 181 tttgtctatcacagagaacattatcacctccctgtggaccctgagacatctgctgtg 240
|||||
Db 37609 TGTGCTATACACAGAGAACATTATACCTCCCTGTGGGACCTGAGACATCTGCTTGTGG 37550
|||||

QY 241 taaatcctgtatgtgtgagcaataacatgaggataaacaccagaccagataccaatgc 300
|||||
Db 37549 TAAATCCTGTATGTGTGAGCAATAACATGAGGATAAACACCAAGTACCCAGAAATCCAAATGC 37490
|||||

QY 301 tgaattttgcttatttccagattcttgcagattcttgcagattcttgcagattt 355
|||||
Db 37489 TGAATATTTGGCTTATTATCCAGATTCTTTGATTGTCAAAGGATTAAATGTT 37435
|||||

RESULT 4
AX106433
LOCUS AX106433 444 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 214 from Patent WO0125272.
AC106433
AC106433.1 GI:13922112
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 444)
XU, J., Skeiky, Y.A., Reed, S.G. and Cheever, M.A.
Compositions and methods for therapy and diagnosis of prostate
cancer
Patent: WO 0125272-A 214 12-APR-2001;
CORIXA CORPORATION (US)
Location/Qualifiers
1. 444
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source
```

```
repeat_region complement(12010..12299)
repeat_region /rpt_family="ALU"
repeat_region complement(12301..13893)
repeat_region /rpt_family="L1"
repeat_region 13897..14096
repeat_region /rpt_family="L1"
repeat_region complement(14777..14838)
repeat_region /rpt_family="L1"
misc_feature 15715..15767
misc_feature /note="similar to human EST T02878 (NID:g3141119)"
misc_feature 19436..19497
misc_feature /note="similar to human EST AA123941 (NID:g1682616)
repeat_region mq22f09.r1"
repeat_region complement(25113..25148)
repeat_region /rpt_family="L1"
repeat_region complement(25561..25578)
repeat_region /rpt_family="L1"
repeat_region complement(26832..27124)
repeat_region /rpt_family="ALU"
misc_feature 27125..27383
misc_feature /note="similar to human EST AA151807 (NID:g1720502)
misc_feature z139g02.r1"
misc_feature 27125..27544
misc_feature /note="similar to human EST AA151796 (NID:g1720491)
misc_feature z139c02.r1"
misc_feature complement(27178..27600)
misc_feature /note="similar to human EST AA149579 (NID:g1720380)
misc_feature z139c02.s1"
misc_feature complement(27441..27890)
misc_feature /note="similar to human EST N52554 (NID:g1193720)
misc_feature yv36a11.s1"
misc_feature 27686..27856
misc_feature /note="similar to human EST W32120 (NID:g1313113)
misc_feature z697c08.r1"
misc_feature 27686..27838
misc_feature /note="similar to human EST W32085 (NID:g1313105)
misc_feature z697b08.r1"
misc_feature complement(27733..28143)
misc_feature /note="similar to human EST N59831 (NID:g1203721)
misc_feature yz77a10.s1"
misc_feature complement(27838..28143)
misc_feature /note="similar to human EST W31561 (NID:g1312680)
misc_feature z697b08.s1"
misc_feature complement(27887..28143)
misc_feature /note="similar to human EST N59830 (NID:g1203720)
misc_feature yz77a09.s1"
misc_feature complement(27997..28144)
misc_feature /note="similar to human EST W31628 (NID:g1312688)
misc_feature z697c08.s1"
misc_feature 28666..28972
misc_feature /note="similar to human EST R09339 (NID:g761262)
misc_feature yf26d05.r1"
misc_feature complement(29587..29853)
misc_feature /note="similar to human EST R09227 (NID:g761150)
misc_feature yf26d05.s1"
misc_feature 30146..30314
misc_feature /note="similar to human EST N78042 (NID:g1240743)
misc_feature yv7le06.r1"
misc_feature complement(31170..31350)
misc_feature /note="similar to human EST N58451 (NID:g1202341)
repeat_region yv7le06.s1"
repeat_region 31970..32006
repeat_region /rpt_family="L1"
repeat_region 32263..32563
repeat_region /rpt_family="ALU"
repeat_region complement(34737..34778)
repeat_region /rpt_family="L1"
repeat_region 36558..36709
repeat_region /rpt_family="ALU"
repeat_region complement(37630..37740)
repeat_region /rpt_family="PTR"
repeat_region 39839..40125
repeat_region /rpt_family="ALU"
```

```
repeat_region 40126..40157
repeat_region /rpt_family="L1"
repeat_region complement(41340..41706)
repeat_region /rpt_family="ALU"
repeat_region complement(41904..41933)
repeat_region /rpt_family="L1"
repeat_region 42348..42409
repeat_region /rpt_family="ALU"
repeat_region complement(42942..42982)
repeat_region /rpt_family="L1"
repeat_region complement(44380..44423)
repeat_region /rpt_family="L1"
repeat_region 44659..44954
repeat_region /rpt_family="ALU"
repeat_region 45422..45448
repeat_region /rpt_family="L1"
repeat_region complement(45615..45669)
repeat_region /rpt_family="L1"
repeat_region complement(45672..45963)
repeat_region /rpt_family="ALU"
repeat_region complement(45964..46597)
repeat_region /rpt_family="L1"
repeat_region 46100..46516
repeat_region /rpt_family="L1"
repeat_region 48370..48785
repeat_region /rpt_family="L1"
repeat_region complement(49577..50216)
repeat_region /rpt_family="L1"
repeat_region complement(50540..50791)
repeat_region /rpt_family="THE"
repeat_region complement(50820..50985)
repeat_region /rpt_family="THE"
repeat_region complement(51124..51156)
repeat_region /rpt_family="L1"
repeat_region 51405..51691
repeat_region /rpt_family="ALU"
repeat_region complement(51721..51764)
repeat_region /rpt_family="L1"
repeat_region complement(51765..52058)
repeat_region /rpt_family="ALU"
repeat_region complement(52252..52467)

Query Match 82.2% Score 405; DB 9; Length 156214;
Best Local Similarity 100.0%; Pred. No. 2.4e-229;
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggctcactgagggtgattggaagtggagatttggccaaatccttgaccattcacttat 60
Db 17445 GGTCACTGTAGGTGTGATTGGAGTGGAGATTTTGGCCAAATCCTTGACCATTGACATTAT 17504

Qy 61 tagatcggtctatcatgtggtcattaggaagtagaataccttaagtgtgttcgaatttt 120
Db 17505 TAGATCGGCTATCATGTGTGTCATAGGAGTAGAATACTTAAGTTTGTCTTCTGAATTTT 17564

Qy 121 tcctcatgtggtgagatgctcactcatcatgaagatgctctcacaaaaataataatt 180
Db 17565 TCCTCATGTGTGTAGATGCTACTCATCATGAAGATGCTCTCACAAAAACAAATATAATAT 17624

Qy 181 tgttgctacacagagaaacattataacctctgtgggacctgagacatctgctgtg99 240
Db 17625 TGTGTGCTATACACAGAGAACATTATACCTCCCTGTGGACCTGAGACATCTGCTGTGGG 17684

Qy 241 taaaatcctgatgtgagcaataacatgaggaataaacacagaccacgaataccaatgc 300
Db 17685 TAAATCCTGATTGATGTGAGCAATAACATGAGGATAAACAGTACCAGATCCAGATCCATGC 17744

Qy 301 tgaattttggcttcattatccagattctttgattgtcaaaaggatttaattgtgtctc 360
Db 17745 TGAATATTTGGCTTCATTATCCAGATTCTTTGATTTGTCAAAAGGATTTAATGTTGTC 17804

Qy 361 agcttgggaccttcagtttaggaacctaaaggatgccagccgaggt 405
Db 17805 AGCTTGGGCACCTTCAGTTAGGACCTTAAGATGCCAGCGCAGGT 17849
```


/code="unnamed protein product"

/codon_start=1

/protein_id="CAC42678.1"

/db_xref="GI:14536714"

/translation="MESISMGSPKSLSETLPNGINGIKDARKVTGVIGSGDFAKS
LTRLRCGVHVIVGSRNPKFASERPHVVDVTHHEDALTKNIIFVAIHRHYTSLW
DLRLHVGKILIDVNNMRINQYPESNAEVLASLFPDSLVIVGFNVSWALQGLPKD
ASQVYICSNIOARQVIELARQLNFIPIIDGLSLSAREIENLPURLFTLMRGPVVV
AISLAFYLYKRRVDVHPYARNOQSDYKIPIEIVNKTLPPIVATLLSLVYLGLL
AAAYOYLTYKTRFPPWLETKWQCRQKQLGSLFFAMVHVAYSLCLPMRNSRYLFL
NMAOQVHANIEWSNEEVRWIEYISFGIMSLGLLSLIATVSIPIVSNALNWRFLS
FIQSTLGYVALLISTFHVLYGWKRAPEEYRYFPNPFVLALVLPISVILLDLQLC
RYPD"

BASE COUNT 648 a 537 c 520 g 748 t
ORIGIN

Query Match 100.0%; Score 493; DB 6; Length 2453;

Best Local Similarity 100.0%; Pred. No. 6e-282;

Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggtcactgtagggtgattggaagtggagattttgccaaatccttgaccattcgacttat 60
|||||
Db 444 GGTCACTGTAGGTGATGTAAGTGGAGATTTTGCCAAATCCTTGACCATTCGACTTAT 503
Qy 61 tagatcggtctatcatgtgctatagaaagtagaaatccttaagtcttctgaattttt 120
|||||
Db 504 TAGATCGCGGTATCATGTGTCATAGGAAGTAGAAATCCTAAGTTTCTCTGAATTTT 563
Qy 121 tcctcattgtgtagatgcactcatcatgaagatgctctcacaaaaacaataatatt 180
|||||
Db 564 TCCFCATGTGTAGATGTCATCATGAGATGCTCTCACAAAAACAATAATATT 623
Qy 181 tgtgtctatacacagagaacattatacctcctgtggacctgagacatctgctgtggg 240
|||||
Db 624 TGTGCTATACACAGACAAATATACCTCCCTGTGGGACCTGAGACATCTGCTGTGGG 683
Qy 241 taaatcctctgattgattgagcaataacaatagaggataaaaccagttaccgaatccaatgc 300
|||||
Db 684 TAAATCCTGATTGATGTAGCAATAACATGAGGATAAAGCAGTACCCAGAAATCCAATGC 743
Qy 301 tgaattattgcttcattattcccaattcttgaattgttcaaggatttaattgtgtctc 360
|||||
Db 744 TGAATATTGGCTTCATATTCCAGATCTTTGATTTGCAAGGATTTAATGTTGTCTC 803
Qy 361 agcttggcacttcagttaggaactaaggatgccagcggcaggtttatatatgcagcaa 420
|||||
Db 804 AGCTTGGGCACTTCAGTTAGGACCTAAGATGCCAGCGGCAGGTTTATATATGCACCAA 863
Qy 421 caattattcaagcgcgacaaaggatttatgaacttgcccgccagttgaatttcattcccat 480
|||||
Db 864 CAATATTCAAGCGCGACAAAGGTTATTGAACCTTGCCCGCAGTTGAATTTCCATTC 923
Qy 481 tgacttggatgcc 493
|||||
Db 924 TGACTTGGGATGCC 936

RESULT 2

HSAC002064

LOCUS

DEFINITION

AC002064

VERSION

AC002064.1

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 156214)

Gattung, S. and Maggi, L.

The sequence of H. sapiens BAC clone RG016J04

Unpublished (1997)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

2 (bases 1 to 156214)

Waterston, R.

Direct Submission

Submitted (09-MAY-1997)

Genome Sequencing Center

Department of Genetics, Washington University

St. Louis, MO 63108, USA

http://genome.wustl.edu/gsc

e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/OTB/CHR7 or send an E-mail to egreen@nhgri.nih.gov

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

VECTOR: pBEO

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of H_RG016J04; actual end is at 156214 of H_RG016J04. The orientation of this clone is unknown.

This clone contains STS SWS2784 (NID:g1113580) and SWS5893 (NID:g454733).

FEATURES

source

1. .156214

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21"

/clone="RG016J04"

/clone_lib="CITB-978SK-B"

15. 40

/rpt_family="L1"

repeat_region

326. 621

/rpt_family="ALU"

repeat_region

complement(977. .1499)

/rpt_family="L1"

repeat_region

3398. 3421

/rpt_family="L1"

repeat_region

5319. 5345

/rpt_family="L1"

repeat_region

complement(8071. .8347)

/rpt_family="ALU"

repeat_region

complement(9406. .9975)

/rpt_family="L1"

repeat_region

complement(10000. .11285)

/rpt_family="L1"

repeat_region

complement(11315. .11984)

/rpt_family="L1"

repeat_region

11666. 11687

/rpt_family="L1"

repeat_region

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 09:50:05 ; Search time 4356.39 Seconds
(without alignments)
2368.195 Million cell updates/sec

Title: US-09-802-520-4

Perfect score: 493

Sequence: 1 ggctactgtagggtgattg.....ttccattgactgggatcc 493

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	493	100.0	2453	6	AXI55249	Sequence
2	405	82.2	156214	9	HSAC002064	Human BAC
3	304	61.7	162928	2	AC104475	Pan trogl
4	209	42.4	444	6	AXI06433	Sequence
5	209	42.4	444	6	AXI40724	Sequence
6	209	42.4	444	6	AXI200584	Sequence
7	209	42.4	444	6	AXI267240	Sequence
8	139	28.2	328	6	AXI06431	Sequence
9	139	28.2	328	6	AXI40722	Sequence
10	139	28.2	328	6	AXI200582	Sequence
11	139	28.2	328	6	AXI267238	Sequence
12	134	27.2	172915	2	AC099742	Papio cyn
13	93	18.9	87401	2	AC021898	Homo sapi
14	28	5.7	205085	2	AC026813	Mus muscu
15	28	5.7	227144	2	AC092404	Mus muscu
16	22	4.5	156882	2	AC023083	Homo sapi
17	22	4.5	169607	2	AC023569	Homo sapi
18	22	4.5	180573	9	AC027689	Homo sapi
19	21	4.3	101743	2	AC091397	Rattus no
20	21	4.3	128615	9	AL591051	Human DNA
21	21	4.3	165686	10	AC090529	Rattus no
22	21	4.3	170278	2	AC020586	Homo sapi
23	20	4.1	42042	3	CEC34B4	278059 Caenorhabdi
24	20	4.1	57220	2	AC100359	Mus muscu
25	20	4.1	70957	2	AP004214	Oryza sat
26	20	4.1	81398	2	AC023821	Homo sapi
27	20	4.1	83921	9	AP001255	Homo sapi
28	20	4.1	89131	2	AC106999	Rattus no
29	20	4.1	95836	9	AC016710	Homo sapi
30	20	4.1	151164	2	AC109808	Homo sapi
31	20	4.1	161086	2	AC097982	Rattus no
32	20	4.1	169109	2	AC025277	Homo sapi
33	20	4.1	174637	2	AC027490	Homo sapi
34	20	4.1	178173	2	AP002414	Homo sapi
35	20	4.1	179497	9	AL355580	Human DNA
36	20	4.1	179691	9	AC006572	Homo sapi
37	20	4.1	182975	9	AP003558	Homo sapi
38	20	4.1	185326	9	AC023600	Homo sapi
39	20	4.1	185619	2	AC095866	Rattus no
40	20	4.1	188596	2	AC027247	Homo sapi
41	20	4.1	188922	2	AC084833	Homo sapi
42	20	4.1	191109	9	AP004249	Homo sapi
43	20	4.1	205402	2	AL645584	Mus muscu
44	20	4.1	224573	9	AC008758	Homo sapi
45	20	4.1	340000	9	AP001686	Homo sapi

ALIGNMENTS

RESULT	1	AXI55249	2453 bp	DNA	linear	PAT 22-JUN-2001
LOCUS	AXI55249	Sequence 7 from Patent WO0140276.				
DEFINITION	AXI55249	AXI55249.1 GI:14536713				
ACCESSION	AXI55249					
VERSION	AXI55249.1	GI:14536713				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1	(bases 1 to 2453)				
AUTHORS	Faris, M. and Jakobovits, A.					
TITLE	Serpentine transmembrane antigens expressed in human prostate cancers and uses thereof.					
JOURNAL	Patent: WO 0140276-A 7 07-JUN-2001;					
FEATURES	Urogenesys, Inc. (US)					
source	Location/Qualifiers					
	1. .2453					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
CDS	355..1719					

THIS PAGE BLANK (USPTO)

	Best Local Similarity	100.0%;	Pred. No. 7.2;		
	Matches	21; Conservative	0; Mismatches	0; Indels	0; Gaps
Qy	173	tgaagagaggaaattggaaaa	193		
Db	165	TGAAGAGAGGAAATTTGAAAAA	185		

RESULT	15				
BE964621		BE964621	1024 bp	mRNA	linear
LOCUS		601658547R1	NIH_MGC_69	Homo sapiens	EST 14-DEC-2000
DEFINITION				clone IMAGE:3885898	3', mRNA sequence.

ACCESSION	BE964621
VERSION	BE964621.2
KEYWORDS	GI:11768128
SOURCE	EST.
	human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1024)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
On Oct 3, 2000 this sequence version replaced gi:10575326.

Contact: Robert Strausberg, Ph.D.
Email: cgabps-r@mail.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

```

FEATURES
source
Location/Qualifiers
1. .1024
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3885898"

```

```

/clone_lib="NIH_MGC_69"
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
244 a 162 c 416 g 202 t

```

	Query Match	4.1%;	Score 21;	DB 10;	Length 1024;
	Best Local Similarity	100.0%;	Pred. No. 7.9;		
	Matches 21;	Conservative	Mismatches	0;	Indels
				0;	Gaps
				0;	
QY	147	ggagggcgcgggggcgcgga	167		
Db	712	GGAGGGCGGGGGCGCGGA	732		

Search completed: September 20, 2002, 08:33:40
Job time: 8704 sec

was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gil4732114[gb]AE129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 223 a 160 c 146 g 183 t
ORIGIN

Query Match 5.6%; Score 29; DB 12; Length 712;
Best Local Similarity 100.0%; Pred. No. 0.00047;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 atctctatgatgggaagccctaagagcct 352
|||||
Db 506 ATCTCTATGATGGGAAGCCCTAAGAGCCT 478

RESULT 12

BF560333
LOCUS

DEFINITION BF560333 336 bp mRNA linear EST 12-DEC-2000
UI-R-EI-fv-a-01-0-UI.r1 UI-R-EI Rattus norvegicus cDNA clone
UI-R-EI-fv-a-01-0-UI 5', mRNA sequence.

ACCESSION BF560333
VERSION BF560333.1 GI:11670063
KEYWORDS EST.

SOURCE

ORGANISM Norway rat.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 336)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene
discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

cDNA Library Preparation: M.B. Soares Lab Clone Distribution:

clones will be available through Research Genetics (www.resgen.com)

This clone is also available through the I.M.A.G.E. Consortium at

LLNL (info@image.llnl.gov). IMAGE ID- 1780077

Seq primer: M13 Forward.

Location/Qualifiers

FEATURES

source

1..336

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-EI-fv-a-01-0-UI"

/clone_lib="UI-R-EI"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not 1; Site_2: Eco RI; The UI-R-EI

library is a subtracted library derived from the UI-R-E0

library. The UI-R-E0 library consisted of a mixture of

individually tagged normalized libraries constructed from

8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-EI) was constructed as follows: PCR

amplified cDNA inserts from a pool of UI-R-E0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-E0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-EI library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

BASE COUNT 82 a 83 c 88 g 83 t
ORIGIN

Query Match 4.8%; Score 25; DB 10; Length 336;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 atctctatgatgggaagccctaaga 348
|||||
Db 102 ATCTCTATGATGGGAAGCCCTAAGA 126

RESULT 13

AL636393
LOCUS

DEFINITION AL636393 653 bp mRNA linear EST 12-DEC-2001
AL636393 XGC-neurula Silurana tropicalis cDNA clone TNeu021f08 5',
mRNA sequence.

ACCESSION AL636393

VERSION AL636393.1 GI:16788372

KEYWORDS EST.

SOURCE western clawed frog.

ORGANISM

Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Silurana.

1 (bases 1 to 653)

Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.

Sanger Xenopus tropicalis EST project 2001 (10_2001)

Unpublished (2001)

Contact: Huckle E

Sanger Centre

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS_SEQUENCE_ID: TNeu021f08.sp6

Sequencing primer: SP6

This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.

Location/Qualifiers

FEATURES

source

1..653

/organism="Silurana tropicalis"

/db_xref="taxon:8364"

/clone="TNeu021f08"

/clone_lib="XGC-neurula"

/dev_stage="neurula"

/lab_host="Escherichia coli DH10B"

/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dT primed from 5ug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

198 a 130 c 143 g 182 t

BASE COUNT

ORIGIN

Query Match 4.3%; Score 22; DB 9; Length 653;
Best Local Similarity 100.0%; Pred. No. 2.1;

LOCUS BB603988 680 bp mRNA linear EST 26-OCT-2001
DEFINITION BB603988 RIKEN full-length enriched, 15 days embryo head Mus musculus cDNA clone D930007L06 5', mRNA sequence.

ACCESSION BB603988
VERSION BB603988.2 GI:16450725
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 680)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Dec 5, 2000 this sequence version replaced gi:11555390. Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,K., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Fanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno H., Fukunishi Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y. Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES
source Location/Qualifiers
1..680
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="D930007L06"
/clone_lib="RIKEN full-length enriched, 15 days embryo head"
/sex="mixed"
/tissue_type="head"
/dev_stage="15 days embryo"
/lab_host="DH10B"
/note="Site:1: Salt; Site:2: BamHI: cDNA library was prepared and sequenced in Mouse Genome Encyclopedia

BASE COUNT 180 a 170 c 168 g 158 t 4 others
ORIGIN

Query Match 5.6%; Score 29; DB 9; Length 680;
Best Local Similarity 100.0%; Pred. No. 0.00046;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 atctctatgatggagagcccttaagagcct 352
|||||
Db 357 ATCTCTATGATGGAGAGCCCTTAGAGCCT 385
|||||

RESULT 11
AZ428664/c 712 bp DNA linear GSS 03-OCT-2000
LOCUS 1M0212F14F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0212F14 F, DNA sequence.

ACCESSION AZ428664
VERSION AZ428664.1 GI:10552677
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 712)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert length: 10000 Std error: 0.00 Plate: 0212 row: F column: 14 Seq primer: CGTTGTAACACGACGCCAGT Class: plasmid ends High quality sequence stop: 712.

FEATURES
source Location/Qualifiers
1..712
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0212F14"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA

ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 (bases 1 to 644)
 Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M., and Hayashizaki, Y.

TITLE
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

JOURNAL
 Unpublished (2001)

COMMENT
 On Dec 5, 2000 this sequence version replaced gi:11555852.
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suihro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., and Hayashizaki, Y.
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
FEATURES
 Location/Qualifiers
 1..644
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="E030001G01"
 /clone_lib="RIKEN full-length enriched, 0 day neonate lung"
 /tissue_type="lung"
 /dev_stage="0 day neonate"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGACGGCGCACTCGAGTTTCTTTTNN 3'], cDNA was prepared by using thermostable thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the

primer adapter of sequence [5', GAGAGAGATTCTCGAGTTAAATTAATATCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC 1."

BASE COUNT
 164 a 164 c 161 g 154 t
ORIGIN

Query Match 5.6%; Score 29; DB 9; Length 644;
 Best Local Similarity 100.0%; Pred. No. 0.00046;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 atctctatgatgggaagccctaagagcct 352
 |||||||||||||||||||||||||||||
 Db 328 ATCTCTATGATGGGAAGCCCTAAGAGCCT 356

RESULT
 9
LOCUS
 BI105662
DEFINITION
 602892429F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5037414 5', mRNA sequence.
ACCESSION
 BI105662
VERSION
 BI105662.1 GI:14556555
KEYWORDS
 EST.
SOURCE
 house mouse.
ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
 1 (bases 1 to 677)
AUTHORS
 NIH-MGC http://mgc.nci.nih.gov/.
TITLE
 National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
 Unpublished (1999)
COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Plate: LLAM1103 row: j column: 07
 High quality sequence stop: 674.

FEATURES
 Location/Qualifiers
 1..677
 /organism="Mus musculus"
 /strain="CZECH II (feral)"
 /db_xref="taxon:10090"
 /clone="IMAGE:5037414"
 /clone_lib="NCI_CGAP_Lu29"
 /tissue_type="spontaneous tumor, metastatic to mammary. Stem cell origin."
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT
 177 a 172 c 160 g 168 t
ORIGIN

Query Match 5.6%; Score 29; DB 10; Length 677;
 Best Local Similarity 100.0%; Pred. No. 0.00046;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 atctctatgatgggaagccctaagagcct 352
 |||||||||||||||||||||||||||||
 Db 210 ATCTCTATGATGGGAAGCCCTAAGAGCCT 238

RESULT
 10
 BB603988


```

Db 333 ATCTCTATGATGGGAAGCCCTAAGAGCCT 361
RESULT 6
LOCUS AI930004
DEFINITION ul60405.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
IMAGE:2123433 5', mRNA sequence.
ACCESSION AI930004
VERSION AI930004
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 506)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:999629
Seq primer: custom primer used.
FEATURES
Location/Qualifiers
1..506
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2123433"
/clone_lib="Sugano mouse kidney mkia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pME18s-FL3; Site_1: DraIII
(CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested
and cloned into distinct DraIII sites of the pME18s-FL3
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTCTGCTCTCTAAAGCTGCG and 3' end
primer CGACCTGCAGCTCGACACA."
BASE COUNT 141 a 124 c 117 g 124 t
ORIGIN
Query Match 5.6%; Score 29; DB 9; Length 506;
Best Local Similarity 100.0%; Pred. NO. 0.00043;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 324 atctctatgagggaagccctaagagcct 352
|||||
Db 231 ATCTCTATGATGGGAAGCCCTAAGAGCCT 259
RESULT 7
LOCUS AI747886
DEFINITION BB604450
LOCUS BB604450
DEFINITION BB604450 RIKEN full-length enriched, 0 day neonate lung Mus
musculus cDNA clone E030001G01 5', mRNA sequence.
ACCESSION BB604450
VERSION BB604450
KEYWORDS EST.
SOURCE house mouse.

```

```

Db 333 ATCTCTATGATGGGAAGCCCTAAGAGCCT 361
RESULT 6
LOCUS AI930004
DEFINITION ul60405.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
IMAGE:2123433 5', mRNA sequence.
ACCESSION AI930004
VERSION AI930004
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 506)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:999629
Seq primer: custom primer used.
FEATURES
Location/Qualifiers
1..506
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2123433"
/clone_lib="Sugano mouse kidney mkia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pME18s-FL3; Site_1: DraIII
(CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested
and cloned into distinct DraIII sites of the pME18s-FL3
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTCTGCTCTCTAAAGCTGCG and 3' end
primer CGACCTGCAGCTCGACACA."
BASE COUNT 141 a 124 c 117 g 124 t
ORIGIN
Query Match 5.6%; Score 29; DB 9; Length 506;
Best Local Similarity 100.0%; Pred. NO. 0.00043;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 324 atctctatgagggaagccctaagagcct 352
|||||
Db 231 ATCTCTATGATGGGAAGCCCTAAGAGCCT 259
RESULT 7
LOCUS AI747886
DEFINITION BB604450
LOCUS BB604450
DEFINITION BB604450 RIKEN full-length enriched, 0 day neonate lung Mus
musculus cDNA clone E030001G01 5', mRNA sequence.
ACCESSION BB604450
VERSION BB604450
KEYWORDS EST.
SOURCE house mouse.

```

```

DEFINITION ul03ell.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
IMAGE:2065004 5', mRNA sequence.
ACCESSION AI747886
VERSION AI747886
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 541)
AUTHORS Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:994192
Seq primer: custom primer used
High quality sequence stop: 498.
FEATURES
Location/Qualifiers
1..541
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2065004"
/clone_lib="Sugano mouse kidney mkia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pME18s-FL3; Site_1: DraIII
(CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCCCTACTGG], digested
and cloned into distinct DraIII sites of the pME18s-FL3
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTCTGCTCTCTAAAGCTGCG and 3' end
primer CGACCTGCAGCTCGACACA."
BASE COUNT 146 a 135 c 126 g 134 t
ORIGIN
Query Match 5.6%; Score 29; DB 9; Length 541;
Best Local Similarity 100.0%; Pred. NO. 0.00043;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 324 atctctatgagggaagccctaagagcct 352
|||||
Db 213 ATCTCTATGATGGGAAGCCCTAAGAGCCT 241
RESULT 8
LOCUS BB604450
DEFINITION BB604450
LOCUS BB604450
DEFINITION BB604450 RIKEN full-length enriched, 0 day neonate lung Mus
musculus cDNA clone E030001G01 5', mRNA sequence.
ACCESSION BB604450
VERSION BB604450
KEYWORDS EST.
SOURCE house mouse.

```

```

BB859054          444 bp  mRNA  linear  EST 26-NOV-2001
LOCUS             BB859054
DEFINITION        BB859054 RIKEN full-length enriched, bladder RCB-0544 MBT-2 CDNA
ACCESSION         BB859054
VERSION           BB859054.1 GI:17100508
KEYWORDS          EST.
SOURCE            house mouse.
ORGANISM          Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 444)
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanaqaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Numasaki,R., Sakazume,N., Sasaki,D., Sato,K.,
Saito,R., Sakai,C., Sakai,K., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
,Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
TITLE             Unpublished (2001)
JOURNAL           Contact: Yoshihide Hayashizaki
COMMENT           Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9226
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES          Location/Qualifiers
source            1. 444
                 /organism="Mus musculus"
                 /db_xref="taxon:10090"
                 /clone="G430005B15"
                 /clone_lib="RIKEN full-length enriched, bladder RCB-0544
                 MBT-2 CDNA"
                 /tissue_type="bladder"
                 /cell_line="RCB-0544 MBT-2"
BASE COUNT       104 a 108 c 141 g 91 t
ORIGIN
Query Match      5.6%; Score 29; DB 9; Length 444;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 324 atctctatgatgggaagccctaagagcct 352
Db 150 ATCTCTATGATGGGAAGCCCTAAGAGCCT 178

BB862648          468 bp  mRNA  linear  EST 26-NOV-2001
LOCUS             BB862648
DEFINITION        BB862648 RIKEN full-length enriched, kidney CCL-142 RAG
musculus CDNA clone G430029C07 5', mRNA sequence.
ACCESSION         BB862648
VERSION           BB862648.1 GI:17104102
KEYWORDS          EST.
SOURCE            house mouse.
ORGANISM          Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 468)
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanaqaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Numasaki,R., Sakazume,N., Sasaki,D., Sato,K.,
Saito,R., Sakai,C., Sakai,K., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
,Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
TITLE             Unpublished (2001)
JOURNAL           Contact: Yoshihide Hayashizaki
COMMENT           Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9226
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES          Location/Qualifiers
source            1. 468
                 /organism="Mus musculus"
                 /db_xref="taxon:10090"
                 /clone="G430029C07"
                 /clone_lib="RIKEN full-length enriched, kidney CCL-142 RAG
                 CDNA"
                 /tissue_type="kidney"
                 /cell_line="CCL-142 RAG"
BASE COUNT       97 a 127 c 149 g 95 t
ORIGIN
Query Match      5.6%; Score 29; DB 9; Length 468;
Best Local Similarity 100.0%; Pred. No. 0.00042;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 324 atctctatgatgggaagccctaagagcct 352

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 06:08:36 ; Search time 5053.42 Seconds
(without alignments)
1380.832 Million cell updates/sec

Title: US-09-802-520-3
Perfect score: 517
Sequence: 1 ggggaagcagctgagtgcg.....aatcctaagtgtgcttctga 517

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : .0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estopl:*
- 7: em_estro:*
- 8: em_hc1:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hc1:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	9.1	483	10	BF883142 QV3-ET021
2	34	6.6	573	10	BM491178 pcp2n.pk0
3	29	5.6	379	9	AI930312 ul63b08.y
4	29	5.6	444	9	BB859054 BB859054
5	29	5.6	468	9	BB862648 BB862648
6	29	5.6	506	9	AI930004 ul60d05.y
7	29	5.6	541	9	AI747886 ul03e11.y
8	29	5.6	644	9	BB604450 BB604450
9	29	5.6	677	10	BI105662 602892429
10	29	5.6	680	9	BB603988 BB603988
11	29	5.6	712	12	Az428664 1M0212F14
12	25	4.8	336	10	BF560333 UI-R-El-f
13	22	4.3	653	9	AL636393 AL636393
14	21	4.1	716	12	AZ976237 2M0251P03
15	21	4.1	1024	10	BE964621 601658547
16	21	4.1	1063	10	BG283322 602407051
17	20	3.9	217	9	BE213064 IpBrn0131

18	20	3.9	481	9	BE211893
19	20	3.9	542	12	AQ619497
20	20	3.9	556	10	BF422234 FMI_13.GO
21	20	3.9	628	10	BJ119688
22	20	3.9	888	9	AL544078
23	19	3.7	245	9	BB183227
24	19	3.7	371	9	AW687319 NF008D01R
25	19	3.7	375	10	W67066 me22b08.r1
26	19	3.7	401	10	R87217
27	19	3.7	402	12	AZ047814
28	19	3.7	409	12	BH060457
29	19	3.7	412	10	H30846
30	19	3.7	430	9	AA121811
31	19	3.7	470	12	AZ519781
32	19	3.7	520	10	BF644068
33	19	3.7	542	10	BG768095
34	19	3.7	550	12	BH262795
35	19	3.7	560	9	BE207123
36	19	3.7	568	10	BF648575
37	19	3.7	568	10	BF650807
38	19	3.7	575	9	AJ388959
39	19	3.7	606	12	AQ688110
40	19	3.7	616	10	BI548782
41	19	3.7	649	10	BF650078
42	19	3.7	654	10	BF648382
43	19	3.7	662	10	BG448710
44	19	3.7	664	10	BF645351
45	19	3.7	675	10	BI264449

ALIGNMENTS

RESULT 1
BF883142/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BF883142 483 bp mRNA linear EST 17-JAN-2001
QV3-ET0211-071200-529-b10 ET0211 Homo sapiens cDNA, mRNA sequence.
BF883142
EST.
BF883142.1 GI:12273268
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 483)
Dias Neto,E., Garcia Correa,R., Vertovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zagó,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=QV3t2-QV3-ET0211-
071200-529-b10&t3=2000-12-07&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 26
High quality sequence stop: 483.
Location/Qualifiers
1. .483
/organism="Homo sapiens"

THIS PAGE BLANK (USPTO)

; LOCATION: 1..4707
US-09-459-066-1

Query Match 3.3%; Score 17; DB 4; Length 4707;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 150 gggcgcgggggcgcg 166
Db 44 GGGCGGGGGGGCGCG 28

Search completed: September 20, 2002, 09:49:23
Job time: 11751 sec

```
Query Match      3.3%; Score 17; DB 4; Length 4707;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17: Conservative 0; Mismatches 0; Indels
```

RESULT 15

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

; TELECOMMUNICATION INFORMATION:

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 170..3077
US-08-447-642-1

```

```

Query Match          3.3%; Score 17; DB 2; Length 4403;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 150 gggcgcgggggggcgcg 166
Db 244 GGGCGCGGGGGCGCG 260

```

```

RESULT 11
US-09-236-503-1
; Sequence 1, Application US/09236503
; Patent No. 6277590
; GENERAL INFORMATION:
; APPLICANT: Barr, Philip J
; TITLE OF INVENTION: Compositions and Methods for PACE 4 and 4.1 Gene and
; FILE REFERENCE: CHIR-009/04US
; CURRENT APPLICATION NUMBER: US/09/236,503
; EARLIER FILING DATE: 1999-01-25
; EARLIER APPLICATION NUMBER: 08/447,642
; EARLIER FILING DATE: 1995-05-23
; EARLIER APPLICATION NUMBER: 08/284,941
; EARLIER FILING DATE: 1994-08-02
; EARLIER APPLICATION NUMBER: 07/848,629
; EARLIER FILING DATE: 1992-03-09
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 1
; LENGTH: 4403
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (170)..(3076)
US-09-236-503-1

```

```

Query Match          3.3%; Score 17; DB 4; Length 4403;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 150 gggcgcgggggggcgcg 166
Db 244 gggcgcgggggggcgcg 260

```

```

RESULT 12
PCT-US93-02147A-1
; Sequence 1, Application PC/TUS9302147A
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
; FILE REFERENCE: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE

```

```

; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02147A
; FILING DATE: 19930309
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,629
; FILING DATE: 09-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: NEELEY PH.D., RICHARD L
; REGISTRATION NUMBER: 30092
; REFERENCE/DOCKET NUMBER: CHIR-009/000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-7622
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (cdna)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 170..3077
PCT-US93-02147A-1

```

```

Query Match          3.3%; Score 17; DB 5; Length 4403;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 150 gggcgcgggggggcgcg 166
Db 244 GGGCGCGGGGGCGCG 260

```

```

RESULT 13
US-09-181-706-1/c
; Sequence 1, Application US/09181706
; Patent No. 6130068
; GENERAL INFORMATION:
; APPLICANT: Melanie K. Spriggs, Michael R. Comeau,
; APPLICANT: Robert F. DuBose, Richard S. Johnson
; TITLE OF INVENTION: VIRAL ENCODED SEMAPHORIN PROTEIN
; FILE REFERENCE: RECEPTOR DNA AND POLYPEPTIDES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Janis C. Henry
; STREET: 51 University St.
; CITY: Seattle
; STATE: WA
; COUNTRY: US
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/181,706
; FILING DATE: October 28, 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```


RESULT 8
US-08-840-204-10
; Sequence 10, Application US/08840204
; Patent No. 6103498
; GENERAL INFORMATION:
; APPLICANT: LAWRENCE, DANIEL A.
; APPLICANT: STEFANSON, STEINGRIMUR P.
; TITLE OF INVENTION: MUTANT PLASMINOGEN ACTIVATOR-INHIBITOR
; TITLE OF INVENTION: TYPE 1 (PAI-1) AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVENUE, NW
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,204
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LIVNAT, SHMUEL
; REGISTRATION NUMBER: 33,949
; REFERENCE/DOCKET NUMBER: 30807-20004.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 822-0168
; TELEX: 90-4030 MRSNFOERSH
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2876 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; ANTI-SENSE: YES
US-08-840-204-10

Query Match 3.3%; Score 17; DB 3; Length 2876;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 gggcgcgggggcgcg 166
|||||
Db 580 GGGCGCGGGGGCGCG 596

RESULT 9
US-08-284-941-1
; Sequence 1, Application US/08284941
; Patent No. 5863756
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,941
; FILING DATE: 2 August 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: NEELEY PH.D., RICHARD L.
; REGISTRATION NUMBER: 30092
; REFERENCE/DOCKET NUMBER: CHIR-009/01US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 843-5070
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 170...3077
US-08-284-941-1

Query Match 3.3%; Score 17; DB 2; Length 4403;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 gggcgcgggggcgcg 166
|||||
Db 244 GGGCGCGGGGGCGCG 260

RESULT 10
US-08-447-642-1
; Sequence 1, Application US/08447642
; Patent No. 5989890
; GENERAL INFORMATION:
; APPLICANT: BARR, PHILIP J
; APPLICANT: KIEFER, MICHAEL C
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PACE 4 AND
; TITLE OF INVENTION: PACE 4.1 GENE AND POLYPEPTIDES IN CELLS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,642
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/284,941
; FILING DATE: 2 August 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: NEELEY PH.D., RICHARD L.
; REGISTRATION NUMBER: 30092
; REFERENCE/DOCKET NUMBER: CHIR-009/01US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 843-5070
; TELEFAX: (415) 857-0663
; TELEX: 380816 COOLEY PA

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/229,515A
FILING DATE: 19 APR 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PERRYMAN, DAVID G
REGISTRATION NUMBER: 33,438
REFERENCE/DOCKET NUMBER: 1414.608
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-688-0770
TELEFAX: 404-688-9880
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-229-515A-13

Query Match 3.3%; Score 17; DB 1; Length 650;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 gggcgcgggggcgcg 166
Db 466 GGGCGCGGGGCGCG 450

RESULT 6
US-08-645-865-13/c
Sequence 13, Application US/08645865
Patent No. 5654406
GENERAL INFORMATION:
APPLICANT: RAZIUDIN
APPLICANT: SARKAR, FAZLUL H
TITLE OF INVENTION: ERBB2 PROMOTER BINDING PROTEIN IN
TITLE OF INVENTION: NEOPLASTIC DISEASE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG PC
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: usa
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/645,865
FILING DATE: 14 MAY 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PERRYMAN, DAVID G
REGISTRATION NUMBER: 33,438
REFERENCE/DOCKET NUMBER: 1414.608
TELEPHONE: 404-688-0770
TELEFAX: 404-688-9880
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-645-865-13

Query Match 3.3%; Score 17; DB 1; Length 650;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 gggcgcgggggcgcg 166
Db 466 GGGCGCGGGGCGCG 450

RESULT 7
US-08-840-204-1/c
Sequence 1, Application US/08840204
Patent No. 6103498
GENERAL INFORMATION:
APPLICANT: LAWRENCE, DANIEL A.
APPLICANT: STEFANSON, STEINGRIMUR P.
TITLE OF INVENTION: MUTANT PLASMINOGEN ACTIVATOR-INHIBITOR
TITLE OF INVENTION: TYPE 1 (PAI-1) AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSEER
STREET: 2000 PENNSYLVANIA AVENUE, NW
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/840,204
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LIVNAT, SHMUEL
REGISTRATION NUMBER: 33,949
REFERENCE/DOCKET NUMBER: 30807-20004.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168
TELEX: 90-4030 MRSNFOERSWSH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2876 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 76..1281
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 145
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 76..144
US-08-840-204-1

Query Match 3.3%; Score 17; DB 3; Length 2876;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 gggcgcgggggcgcg 166
Db 2297 GGGCGCGGGGCGCG 2281

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,090
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jean M. Silveri
; REGISTRATION NUMBER: 39,030
; REFERENCE/DOCKET NUMBER: MNI-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2689 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 291..1625
; US-08-985-090-1

Query Match 3.7%; Score 19; DB 2; Length 2689;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 gccacggcagccaccctgc 42
|||||
Db 1759 GCCACGGCAGCCACCCTGC 1777

RESULT 3
US-09-165-543-1
; Sequence 1, Application US/09165543
; Patent No. 6093545
; GENERAL INFORMATION:
; APPLICANT: Andrew D.J. Goodearl and Sandra Glucksmann
; TITLE OF INVENTION: Muscarinic Receptors and Uses Therefor
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/165,543
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/042,780
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth A. Hanley
; REGISTRATION NUMBER: 33,505
; REFERENCE/DOCKET NUMBER: MNI-032CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2689 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 291..1625
; US-09-165-543-1

Query Match 3.7%; Score 19; DB 3; Length 2689;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 gccacggcagccaccctgc 42
|||||
Db 1759 GCCACGGCAGCCACCCTGC 1777

RESULT 4
US-09-167-354-5
; Sequence 5, Application US/09167354A
; Patent No. 6136559
; GENERAL INFORMATION:
; APPLICANT: Lovenberg, Timothy
; APPLICANT: Erlander, Mark
; APPLICANT: Pivati, Jayashree
; APPLICANT: Huvar, Arne
; TITLE OF INVENTION: DNA ENCODING A HUMAN HISTAMINE RECEPTOR OF THE H3
; TITLE OF INVENTION: SUBTYPE
; FILE REFERENCE: JMW
; CURRENT APPLICATION NUMBER: US/09/167,354A
; CURRENT FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 2699
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: CDNA
; US-09-167-354-5

Query Match 3.7%; Score 19; DB 3; Length 2699;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 gccacggcagccaccctgc 42
|||||
Db 1767 gccacggcagccaccctgc 1785

RESULT 5
US-08-229-515A-13/c
; Sequence 13, Application US/08229515A
; Patent No. 5518885
; GENERAL INFORMATION:
; APPLICANT: RAZIUDIN
; APPLICANT: SAKKAR, FAZLUL H
; TITLE OF INVENTION: ERBB2 PROMOTER BINDING PROTEIN IN
; TITLE OF INVENTION: NEOPlastic DISEASE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG PC
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: usa
; ZIP: 30303

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 06:33:32 ; Search time 139.75 Seconds
(without alignments)
908.712 Million cell updates/sec

Title: US-09-802-520-3
Perfect score: 517
Sequence: 1 ggggaagcagctggagtgcg.....aatcctaagtgtgcttctga 517

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	3.7	375	US-09-167-354-8	Sequence 8, Appli
2	19	3.7	2689	US-08-985-090-1	Sequence 1, Appli
3	19	3.7	2689	US-09-165-543-1	Sequence 1, Appli
4	19	3.7	2699	US-09-167-354-5	Sequence 5, Appli
C 5	17	3.3	650	US-08-229-515A-13	Sequence 13, Appli
C 6	17	3.3	650	US-08-645-865-13	Sequence 13, Appli
C 7	17	3.3	2876	US-08-840-204-1	Sequence 1, Appli
8	17	3.3	2876	US-08-840-204-10	Sequence 10, Appli
9	17	3.3	4403	US-08-284-941-1	Sequence 1, Appli
10	17	3.3	4403	US-08-447-642-1	Sequence 1, Appli
11	17	3.3	4403	US-09-236-503-1	Sequence 1, Appli
12	17	3.3	4403	PCT-US93-02147A-1	Sequence 1, Appli
C 13	17	3.3	4707	US-09-181-706-1	Sequence 1, Appli
C 14	17	3.3	4707	US-09-458-791-1	Sequence 1, Appli
C 15	17	3.3	4707	US-09-459-066-1	Sequence 1, Appli
C 16	17	3.3	12284	US-08-876-991-1	Sequence 1, Appli
C 17	17	3.3	12284	US-09-059-853-1	Sequence 1, Appli
18	16	3.1	1240	US-08-869-506-1	Sequence 1, Appli
19	16	3.1	1240	US-09-128-967-1	Sequence 1, Appli
C 20	16	3.1	1981	US-08-747-574-1	Sequence 1, Appli
21	16	3.1	2154	US-09-488-856A-3	Sequence 3, Appli
C 22	16	3.1	2664	US-08-942-819-1	Sequence 1, Appli
C 23	16	3.1	3879	US-08-530-010-1	Sequence 1, Appli
C 24	16	3.1	3879	US-08-484-101B-1	Sequence 1, Appli
C 25	16	3.1	3879	US-08-714-524D-1	Sequence 1, Appli
26	16	3.1	5176	US-09-182-024A-1	Sequence 1, Appli
C 27	16	3.1	5643	US-09-079-415-5	Sequence 5, Appli

C 28	16	3.1	5643	3	US-08-750-458A-1	Sequence 1, Appli
29	16	3.1	6027	2	US-08-968-542C-1	Sequence 1, Appli
30	16	3.1	12394	4	US-09-488-856A-10	Sequence 10, Appli
31	16	3.1	12912	2	US-08-460-751-1	Sequence 1, Appli
32	16	3.1	14060	3	US-08-658-136-4	Sequence 4, Appli
C 33	16	3.1	28804	2	US-08-592-874-1	Sequence 1, Appli
C 34	16	3.1	28804	3	US-09-096-942-2	Sequence 2, Appli
C 35	16	3.1	28804	3	US-09-096-867-2	Sequence 2, Appli
36	16	3.1	31571	1	US-08-323-443B-1	Sequence 1, Appli
37	16	3.1	53526	3	US-08-658-136-2	Sequence 1, Appli
38	16	3.1	53577	3	US-08-658-136-1	Sequence 1, Appli
39	15	2.9	32	2	US-08-997-080-129	Sequence 129, App
40	15	2.9	32	2	US-08-997-362-129	Sequence 129, App
41	15	2.9	32	4	US-09-095-855-129	Sequence 129, App
42	15	2.9	32	4	US-09-324-542-129	Sequence 129, App
C 43	15	2.9	73	1	US-07-626-618A-11	Sequence 11, Appli
C 44	15	2.9	73	1	US-08-333-977-11	Sequence 11, Appli
C 45	15	2.9	105	2	US-08-623-471-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-167-354-8
; Sequence 8, Application US/09167354A
; Patent No. 6136559
; GENERAL INFORMATION:
; APPLICANT: Lovenberg, Timothy
; APPLICANT: Erlander, Mark
; APPLICANT: Pyati, Jayashree
; APPLICANT: Huvar, Arne
; TITLE OF INVENTION: DNA ENCODING A HUMAN HISTAMINE RECEPTOR OF THE H3
; TITLE OF INVENTION: SUBTYPE
; FILE REFERENCE: JWV
; CURRENT APPLICATION NUMBER: US/09/167,354A
; CURRENT FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:CDNA
US-09-167-354-8

Query Match 3.7%; Score 19; DB 3; Length 375;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 gccacggcagccaccctgc 42
Db 295 gccacggcagccaccctgc 313

RESULT 2
US-08-985-090-1
; Sequence 1, Application US/08985090
; Patent No. 5885893
; GENERAL INFORMATION:
; APPLICANT: Andrew D. J. Goodearl
; TITLE OF INVENTION: MUSCARINIC RECEPTORS AND USES THEREFOR
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESS: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2192 BP; 559 A; 507 C; 551 G; 575 T; 0 Other;

Query Match 34.4%; Score 178; DB 23; Length 2192;
Best Local Similarity 99.6%; Pred. NO. 6.3e-85;
Matches 228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 280 aggatattctgtgtagcttgggaagtcgcgtatcatcgaatcaatctctatgatggaa 339

Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 260 aggatattctgtgtagcttgggaagtcgcgtatcatcgaatcaatctctatgatggaa 319

QY 340 gccctaagagccttagtgaaactgtttacctaataatggcataaaatggatcaaaagatgcaa 399

Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 320 gccctaagagccttagtgaaactttttacctaataatggcataaaatggatcaaaagatgcaa 379

QY 400 ggaagtcactgtaggtgtgattggaagtggagattttgccaaatccttgaccattcgac 459

Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 380 ggaagtcactgtaggtgtgattggaagtggagattttgccaaatccttgaccattcgac 439

QY 460 ttattagatgcgcctatcatgtgtcataggaagttagaaatcctaagtt 508

Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 440 ttattagatgcgcctatcatgtgtcataggaagttagaaatcctaagtt 488

Search completed: September 20, 2002, 09:59:43
Job time: 11701 sec

|||||
Db 403 gccctaagagccttagtgaaacttttttacctaataatgacataaaatgggtatcaaaagatgcaa 462
Qy 400 ggaaggtcactgttaggtgtgattggaagtgagatgttttgcacaaatccttgaccattcgac 459
Db 463 ggaaggtcactgttaggtgtgattggaagtgagatgttttgcacaaatccttgaccattcgac 522
Qy 460 ttattagatcggtctatcatgtgctataggaagtagaataatcctaagtt 508
Db 523 ttattagatcggtctatcatgtgctataggaagtagaataatcctaagtt 571
RESULT 14
AAI04420
ID AAI04420 standard; DNA; 1953 BP.
XX
AC AAI04420;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #4411 used to measure gene expression in human breast sample.
XX
KW Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
XX Claim 25; SEQ ID No 4411; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;

Query Match 34.4%; Score 178; DB 22; Length 1953;
Best Local Similarity 99.6%; Pred. No. 6.3e-85;
Matches 228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 280 aggatattcttggtgatcttggaaagtgccgtatcatggaatcaatctctatgatggaa 339
Db 343 aggatattcttggtgatcttggaaagtgccgtatcatggaatcaatctctatgatggaa 402
Qy 340 gccctaagagccttagtgaaacttttttacctaataatgacataaaatgggtatcaaaagatgcaa 399
Db 403 gccctaagagccttagtgaaacttttttacctaataatgacataaaatgggtatcaaaagatgcaa 462
Qy 400 ggaaggtcactgttaggtgtgattggaagtgagatgttttgcacaaatccttgaccattcgac 459
Db 463 ggaaggtcactgttaggtgtgattggaagtgagatgttttgcacaaatccttgaccattcgac 522
Qy 460 ttattagatcggtctatcatgtgctataggaagtagaataatcctaagtt 508
Db 523 ttattagatcggtctatcatgtgctataggaagtagaataatcctaagtt 571
RESULT 15
AAS76493
ID AAS76493 standard; cDNA; 2192 BP.
XX
AC AAS76493;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #12297.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG12306.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 12297; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.

```
QY 340 gcctaagagccttagtgaaactgtttacctaataatggcataaaatggtatcaaaagatgcaa 399
Db 403 gcctaagagccttagtgaaactgtttacctaataatggcataaaatggtatcaaaagatgcaa 462
QY 400 ggaagtcacttagtgatggagtggaagtggaattttgccaaatccttgaccattcgac 459
Db 463 ggaagtcacttagtgatggagtggaagtggaattttgccaaatccttgaccattcgac 522
QY 460 ttattagatcgcgctcatcatgtggtcataggaagttagaataatcctaagtt 508
Db 523 ttattagatcgcgctcatcatgtggtcataggaagttagaataatcctaagtt 571

RESULT 12
AAI14606
ID AAI14606 standard; DNA; 1953 BP.
XX
AC AAI14606;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #4539 for gene expression analysis in human cervical cell sample.
XX
KW Probe: human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US006070.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 4539; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;
```

```
Query Match 34.4%; Score 178; DB 22; Length 1953;
Best Local Similarity 99.6%; Pred. No. 6.3e-85;
Matches 228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 280 aggatattcttggtgatcttggaagtgccgtatcatggaataatcctctatgatggaa 339
Db 343 aggatattcttggtgatcttggaagtgccgtatcatggaataatcctctatgatggaa 402
QY 340 gcctaagagccttagtgaaactgtttacctaataatggcataaaatggtatcaaaagatgcaa 399
```

```
Db 343 aggatattcttggtgatcttggaagtgccgtatcatggaataatcctctatgatggaa 402
QY 340 gcctaagagccttagtgaaactgtttacctaataatggcataaaatggtatcaaaagatgcaa 399
Db 403 gcctaagagccttagtgaaactgtttacctaataatggcataaaatggtatcaaaagatgcaa 462
QY 400 ggaagtcacttagtgatggagtggaagtggaattttgccaaatccttgaccattcgac 459
Db 463 ggaagtcacttagtgatggagtggaagtggaattttgccaaatccttgaccattcgac 522
QY 460 ttattagatcgcgctcatcatgtggtcataggaagttagaataatcctaagtt 508
Db 523 ttattagatcgcgctcatcatgtggtcataggaagttagaataatcctaagtt 571

RESULT 13
AAI35978
ID AAI35978 standard; DNA; 1953 BP.
XX
AC AAI35978;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #4664 used to measure gene expression in human placenta sample.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US006663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488997/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 4664; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;
```

```
Query Match 34.4%; Score 178; DB 22; Length 1953;
Best Local Similarity 99.6%; Pred. No. 6.3e-85;
Matches 228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 280 aggatattcttggtgatcttggaagtgccgtatcatggaataatcctctatgatggaa 339
Db 343 aggatattcttggtgatcttggaagtgccgtatcatggaataatcctctatgatggaa 402
QY 340 gcctaagagccttagtgaaactgtttacctaataatggcataaaatggtatcaaaagatgcaa 399
```


QY 400 ggaaggtcactgtaggtgtgattggaagtggagattttgccaatactctgaccattcgac 459
|||||
Db 463 ggaaggtcactgtaggtgtgattggaagtggagattttgccaatactctgaccattcgac 522
|||||
QY 460 ttattagatcgcgctcatcatgtgtgtcataggaagtgaatactctaagtt 508
|||||
Db 523 ttattagatcgcgctcatcatgtgtgtcataggaagtgaatactctaagtt 571
|||||

RESULT 8
ABA56335.
ID ABA56335 standard; DNA; 1953 BP.
XX AC ABA56335;
XX
DT 01-FEB-2002 (first entry)
DE Human foetal liver single exon nucleic acid probe #4640.
XX
DE Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
KW Homo sapiens.
OS WO200157277-A2.
XX
PN 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 1; SEQ ID NO 4640; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO.int/pub/published_pct_sequences.
XX
XX Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;

Query Match 34.4%; Score 178; DB 22; Length 1953;
Best Local Similarity 99.6%; Pred. No. 6.3e-85;
Matches 228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 280 aggatattcttgatcttggaagtgtccgtatcatggaatactctatgatggaa 339
|||||
Db 343 aggatattcttgatcttggaagtgtccgtatcatggaatactctatgatggaa 402
|||||
QY 340 gccctaagaccttagtgaacctgtttacctaattggcctaataatggtatcaaatgacaa 399
|||||
Db 403 gccctaagaccttagtgaacctgtttacctaattggcctaataatggtatcaaatgacaa 462
|||||

QY 400 ggaaggtcactgtaggtgtgattggaagtggagattttgccaatactctgaccattcgac 459
|||||
Db 463 ggaaggtcactgtaggtgtgattggaagtggagattttgccaatactctgaccattcgac 522
|||||
QY 460 ttattagatcgcgctcatcatgtgtgtcataggaagtgaatactctaagtt 508
|||||
Db 523 ttattagatcgcgctcatcatgtgtgtcataggaagtgaatactctaagtt 571
|||||

RESULT 9
ABA25976
ID ABA25976 standard; DNA; 1953 BP.
XX AC ABA25976;
XX
DT 23-JAN-2002 (first entry)
DE Probe #4442 for gene expression analysis in human heart cell sample.
XX
DE Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488999/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
XX Claim 1; SEQ ID No 4442; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;

Query Match 34.4%; Score 178; DB 22; Length 1953;
Best Local Similarity 99.6%; Pred. No. 6.3e-85;
Matches 228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 280 aggatattcttgatcttggaagtgtccgtatcatggaatactctatgatggaa 339
|||||

KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
 KW liver cancer; lung cancer; cytostatic; ss.
 XX Homo sapiens.

XX Key Location/Qualifiers
 XX CDS 178..1650
 FT /*tag= a
 FT /product= "STMPL"

XX WO200172962-A2.

XX 04-OCT-2001.

XX 23-MAR-2001; 2001WO-US09410.

XX 24-MAR-2000; 2000US-191929P.

XX (SAAT/) SAATCIOGLU F.

XX Saatioglu F;

XX WPI; 2001-662926/76.

XX P-PSDB; AAU10187.

XX New polynucleotide for the diagnosis, prevention and treatment for
 PT prostate and testis disorders, particularly prostate cancer, comprises
 PT prostate-specific or testis-specific nucleic acids -

XX Claim 5; Fig 4D; 114pp; English.

XX The invention relates to substantially pure prostate-specific or
 CC testis-specific polypeptides and the nucleic acids encoding them.
 CC Also included are vectors and host cells expressing the proteins, a
 CC transgenic animal expressing the protein, antibodies against the
 CC proteins, probes for detecting the nucleic acids, antisense molecules
 CC for the nucleic acids and methods of isolating modulators of the
 CC proteins. Compounds that modulate the prostate specific or testis
 CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,
 CC cryptorchidism, undescended, retractile, ascending or vanished
 CC testis. Other proliferative disorders for which the modulators may be
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The
 CC present sequence represents the open reading frame of a prostate
 CC specific protein, Six-Transmembrane Protein of Prostate 1, STMPL.

XX Sequence 4329 BP; 1315 A; 817 C; 790 G; 1407 T; 0 other;

Query Match 44.38; Score 229; DB 22; Length 4329;
 Best Local Similarity 100.0%; Pred. No. 3.5e-112;
 Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 280 aggatattcttggtgattggaagtcgcgtatcatggaatcaatctctatgatggaa 339
 |||||||

Db 143 aggatattcttggtgattggaagtcgcgtatcatggaatcaatctctatgatggaa 202
 |||||||

Qy 340 gccctaagagccttagtgaaactgtttacctaataatggcataaaatggatcaaatgcaa 399
 |||||||

Db 203 gccctaagagccttagtgaaactgtttacctaataatggcataaaatggatcaaatgcaa 262
 |||||||

Qy 400 ggaaggtcaactgtagggtgattggaagtggaagttttgcacaaatccttgaccattgac 459
 |||||||

Db 263 ggaaggtcaactgtagggtgattggaagtggaagttttgcacaaatccttgaccattgac 322
 |||||||

Qy 460 ttattagatgcggtatcatgtgtcataggaagtagaatacctaagtt 508
 |||||||

Db 323 ttattagatgcggtatcatgtgtcataggaagtagaatacctaagtt 371
 |||||||

RESULT 7

ABA45820
 ID ABA45820 standard; DNA; 1953 BP.

XX ABA45820;

AC ABA45820;

XX 01-FEB-2002 (first entry)

DT Human breast cell single exon nucleic acid probe #4515.

DE Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.

XX Homo sapiens.

XX WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00662.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -

XX Claim 1; SEQ ID NO 4515; 327pp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;

Query Match 34.48; Score 178; DB 22; Length 1953;
 Best Local Similarity 99.68; Pred. No. 6.3e-85;
 Matches 228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 280 aggatattcttggtgattggaagtcgcgtatcatggaatcaatctctatgatggaa 339
 |||||||

Db 343 aggatattcttggtgattggaagtcgcgtatcatggaatcaatctctatgatggaa 402
 |||||||

Qy 340 gccctaagagccttagtgaaactgtttacctaataatggcataaaatggatcaaatgcaa 399
 |||||||

Db 403 gccctaagagccttagtgaaactgtttacctaataatggcataaaatggatcaaatgcaa 462
 |||||||

DE Human DNA for Six-Transmembrane Protein of Prostate 1, STMP1.
 XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
 KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
 KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
 KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
 KW liver cancer; lung cancer; cytostatic; ds.
 XX Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 89
 FT /*tag= a
 FT /note= "Represents 338 nucleotides of intron 1"
 FT 162
 FT misc_feature
 FT /*tag= b
 FT /note= "Represents 12713 nucleotides of intron 2"
 FT 200..1702
 FT CDS
 FT /*tag= c
 FT /product= "STMP1"
 FT 697
 FT misc_feature
 FT /*tag= d
 FT /note= "Represents 1396 nucleotides of intron 3"
 FT 1225
 FT misc_feature
 FT /*tag= e
 FT /note= "Represents 2372 nucleotides of intron 4"
 FT 1410
 FT misc_feature
 FT /*tag= f
 FT /note= "Represents 2299 nucleotides of intron 5"
 FT
 XX WO200172962-A2.
 PN
 XX
 PD 04-OCT-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09410.
 XX
 PR 24-MAR-2000; 2000US-191929P.
 XX
 PA (SAAT/) SAATCIOGLU F.
 PI Saatcioglu F;
 XX WPI; 2001-662926/76.
 DR P-PSDB; AAU10187.
 XX
 PT New polynucleotide for the diagnosis, prevention and treatment for
 PT prostate and testis disorders, particularly prostate cancer, comprises
 PT prostate-specific or testis-specific nucleic acids -
 XX
 PS Example 3; Fig 4B; 114pp; English.
 XX
 CC The invention relates to substantially pure prostate-specific or
 CC testis-specific polypeptides and the nucleic acids encoding them.
 CC Also included are vectors and host cells expressing the proteins, a
 CC transgenic animal expressing the protein, antibodies against the
 CC proteins, probes for detecting the nucleic acids, antisense molecules
 CC for the nucleic acids and methods of isolating modulators of the
 CC proteins. Compounds that modulate the prostate specific or testis
 CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,
 CC cryptorchidism, undescended, retractile, ascending or vanished
 CC testis. Other proliferative disorders for which the modulators may be
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The
 CC present sequence encodes a prostate specific protein, Six-Transmembrane
 CC Protein of Prostate 1, STMP1.
 XX
 SQ Sequence 1725 BP; 476 A; 340 C; 387 G; 517 T; 5 other;

Query Match 44.3%; Score 229; DB 22; Length 1725;
 Best Local Similarity 100.0%; Pred. No. 3.4e-112;

Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 280 aggatattcttggtgattggaagtgcctgatacatggaatcaatctctatgatggaa 339
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 165 aggatattcttggtgattggaagtgcctgatacatggaatcaatctctatgatggaa 224
 QY 340 gccctaagagccttagtgaaactgtttacctaataatggcataaaatggtatcaaatgcaa 399
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 225 gccctaagagccttagtgaaactgtttacctaataatggcataaaatggtatcaaatgcaa 284
 QY 400 ggaaggtcaactgttagtgattggaagtggagattttgcaaatctctgaccattggac 459
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 285 ggaaggtcaactgttagtgattggaagtggagattttgcaaatctctgaccattggac 344
 QY 460 ttattagatcggtcatcatggtcatggaagttagaataatccttaagt 508
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 345 ttattagatcggtcatcatggtcatggaagttagaataatccttaagt 393
 RESULT 4
 AAS15813
 ID AAS15813 standard; CDNA; 2102 BP.
 XX
 AC AAS15813;
 XX
 DT 16-JAN-2002 (first entry)
 XX
 DE Human cDNA encoding ORF3 of Six-Transmembrane Protein of Prostate 1.
 XX
 KW Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
 KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
 KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
 KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
 KW liver cancer; lung cancer; cytostatic; ss; ORF3.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS
 FT 171..1430
 FT /*tag= a
 FT /product= "STMP1, ORF3"
 FT
 XX WO200172962-A2.
 PN
 XX
 PD 04-OCT-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09410.
 XX
 PR 24-MAR-2000; 2000US-191929P.
 XX
 PA (SAAT/) SAATCIOGLU F.
 XX Saatcioglu F;
 PI WPI; 2001-662926/76.
 DR P-PSDB; AAU10189.
 XX
 PT New polynucleotide for the diagnosis, prevention and treatment for
 PT prostate and testis disorders, particularly prostate cancer, comprises
 PT prostate-specific or testis-specific nucleic acids -
 XX
 PS Claim 4; Fig 4K; 114pp; English.
 XX
 CC The invention relates to substantially pure prostate-specific or
 CC testis-specific polypeptides and the nucleic acids encoding them.
 CC Also included are vectors and host cells expressing the proteins, a
 CC transgenic animal expressing the protein, antibodies against the
 CC proteins, probes for detecting the nucleic acids, antisense molecules
 CC for the nucleic acids and methods of isolating modulators of the
 CC proteins. Compounds that modulate the prostate specific or testis
 CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,

```
XX
PI Saatcioglu F;
XX
DR WPI: 2001-662926/76.
DR P-PSDB; AAU10188.
XX
XX New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids -
XX
XX Claim 4; Fig 4H; 114pp; English.
PS
XX
CC The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence encodes prostate specific protein, Six-Transmembrane
CC Protein of Prostate 1, STMP1, ORF2.
XX
XX Sequence 1561 BP; 420 A; 319 C; 346 G; 476 T; 0 other;
SQ

Query Match 44.3%; Score 229; DB 22; Length 1561;
Best Local Similarity 100.0%; Pred. No. 3.4e-112;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 aggatattctgtgacgttggaagtcgcgtatcatggaatcaatctctatgatggaa 339
Db 153 aggatattctgtgacgttggaagtcgcgtatcatggaatcaatctctatgatggaa 212

QY 340 gccctaagagccttagtgaaactgtttacctaataatggcataaaatggtatcaaatgcaa 399
Db 213 gccctaagagccttagtgaaactgtttacctaataatggcataaaatggtatcaaatgcaa 272

QY 400 ggaagtcactgtagtgatgagtggaagtggaagatttgcacaaatccttgaccattcgac 459
Db 273 ggaagtcactgtagtgatgagtggaagtggaagatttgcacaaatccttgaccattcgac 332

QY 460 ttattagatcgcgctatcatgtggtcataggaagtagaataatcctaaagtt 508
Db 333 ttattagatcgcgctatcatgtggtcataggaagtagaataatcctaaagtt 381

RESULT 2
AAS15802
ID AAS15802 standard; cDNA; 1680 BP.
XX
AC AAS15802;
XX
XX 16-JAN-2002 (first entry)
DT
XX Human cDNA encoding Six-Transmembrane Protein of Prostate 1, STMP1.
DE
XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 178..1650
FT CDS
```

```
FT
XX
XX
PN WO200172962-A2.
XX
PD 04-OCT-2001.
XX
XX 23-MAR-2001; 2001WO-US09410.
PF
XX
XX 24-MAR-2000; 2000US-191929P.
PR
XX
XX (SAAT/) SAATCIOGLU F.
PA
XX
XX Saatcioglu F;
PI
XX
XX WPI: 2001-662926/76.
DR P-PSDB; AAU10187.
DR
XX
XX New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids -
XX
XX Claim 4; Fig 4E; 114pp; English.
PS
XX
XX The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence encodes a prostate specific protein, Six-Transmembrane
CC Protein of Prostate 1, STMP1.
XX
XX Sequence 1680 BP; 467 A; 334 C; 373 G; 506 T; 0 other;
SQ

Query Match 44.3%; Score 229; DB 22; Length 1680;
Best Local Similarity 100.0%; Pred. No. 3.4e-112;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 aggatattctgtgacgttggaagtcgcgtatcatggaatcaatctctatgatggaa 339
Db 143 aggatattctgtgacgttggaagtcgcgtatcatggaatcaatctctatgatggaa 202

QY 340 gccctaagagccttagtgaaactgtttacctaataatggcataaaatggtatcaaatgcaa 399
Db 203 gccctaagagccttagtgaaactgtttacctaataatggcataaaatggtatcaaatgcaa 262

QY 400 ggaagtcactgtagtgatgagtggaagtggaagatttgcacaaatccttgaccattcgac 459
Db 263 ggaagtcactgtagtgatgagtggaagtggaagatttgcacaaatccttgaccattcgac 322

QY 460 ttattagatcgcgctatcatgtggtcataggaagtagaataatcctaaagtt 508
Db 323 ttattagatcgcgctatcatgtggtcataggaagtagaataatcctaaagtt 371

RESULT 3
AAS15793
ID AAS15793 standard; DNA; 1725 BP.
XX
AC AAS15793;
XX
XX 16-JAN-2002 (first entry)
DT
XX
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 06:44:42 ; Search time 600.75 seconds
(without alignments)
1477.561 Million cell updates/sec

Title: US-09-802-520-3
Perfect score: 517
Sequence: 1 ggggaagcagctggagtcgcg.....aatcctaagtggctctga 517

Scoring table: OLIGO_NUC
Gapop 60.0 , Capext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	229	44.3	1561	22 AAS15811	Human cDNA encodin
2	229	44.3	1680	22 AAS15802	Human cDNA encodin
3	229	44.3	1725	22 AAS15793	Human DNA for Six-
4	229	44.3	2102	22 AAS15813	Human cDNA encodin
5	229	44.3	2453	22 AAD07072	Human six transmem
6	229	44.3	4329	22 AAS15801	Human ORF of Six-T
7	178	34.4	1953	22 ABA45820	Human breast cell
8	178	34.4	1953	22 ABA56335	Human foetal liver
9	178	34.4	1953	22 ABA25976	Probe #4442 for ge

10	178	34.4	1953	22	AAK04514	Human brain expres
11	178	34.4	1953	22	AAK30016	Human bone marrow
12	178	34.4	1953	22	AAI14606	Probe #4539 for ge
13	178	34.4	1953	22	AAI35978	Probe #4664 used t
14	178	34.4	1953	22	AAI04420	Probe #4411 used t
15	178	34.4	2192	23	AAI76493	DNA encoding novel
16	178	34.4	2238	22	AAI5810	Human ORF2 of Six-
17	178	34.4	3900	23	AAI5810	DNA encoding novel
18	176	34.0	525	22	AAI5796	Human Six-Transmem
19	176	34.0	525	22	AAI5805	Human Six-Transmem
20	176	34.0	539	22	ABA50938	Human breast cell
21	176	34.0	539	22	ABA68917	Human foetal liver
22	176	34.0	539	22	ABA35870	Probe #14336 for g
23	176	34.0	539	22	AAK17249	Human brain expres
24	176	34.0	539	22	AAK43038	Human bone marrow
25	176	34.0	539	22	AAI23801	Probe #13734 for g
26	176	34.0	539	22	AAI49113	Probe #17799 used
27	176	34.0	539	22	AAI09410	Probe #9401 used t
28	100	19.3	3185	23	AAI51627	DNA encoding novel
29	31	6.0	83	22	AAI51794	Human Six-Transmem
30	30	5.8	75	22	AAI515803	Human Six-Transmem
31	20	3.9	1651	24	ABL34327	Human immune syste
32	19	3.7	375	21	AAI09065	Human histamine H
33	19	3.7	375	22	AAI52367	Nucleotide sequenc
34	19	3.7	739	22	AAI55664	Human neuroblastom
35	19	3.7	894	21	AAI77127	Human ORFX ORF2682
36	19	3.7	1893	20	AAI584582	G-protein conjugat
37	19	3.7	1893	21	AAI70648	Human G-protein co
38	19	3.7	1902	22	AAI94219	Human full-length
39	19	3.7	2173	24	AAI562391	CDNA sequence #178
40	19	3.7	2576	23	AAI56867	DNA encoding novel
41	19	3.7	2689	20	AAI44572	Human muscarinic a
42	19	3.7	2689	20	AAI59167	Human G protein co
43	19	3.7	2689	20	AAI02885	Human mAHR-6 cDNA
44	19	3.7	2699	21	AAI09061	Human histamine H3
45	19	3.7	2699	22	AAI62365	Complete nucleotid

ALIGNMENTS

RESULT 1

AAS15811 standard; cDNA; 1561 BP.

ID AAS15811;

AC AAS15811;

XX 16-JAN-2002 (first entry)

DT Human cDNA encoding ORF2 of Six-Transmembrane Protein of Prostate 1.

XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
DE benign prostatic hyperplasia; acute prostatitis; testicular cancer;
DE cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; ss; ORF2.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 188..1552

FT FT /*tag= a

FT FT /product= "STMP1, ORF2"

XX WO200172962-A2.

XX 04-OCT-2001.

XX 23-MAR-2001; 2001WO-US09410.

XX 24-MAR-2000; 2000US-191929P.

XX (SAAT/) SAATCIOGLU F.

THIS PAGE BLANK (USPTO)

```
repeat_region 130..146 /note="alpha element"
BASE COUNT 30 a /note="alpha element" 15 t
ORIGIN 2 bp upstream of Nari site.

Query Match 3.9%; Score 20; DB 14; Length 210;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 152 gcgcgggggcgcgcggagaaa 171
|||||
Db 132 GCGCGGGGGCGCGGAGAAA 151

RESULT 15
AX347229
LOCUS AX347229 1651 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 2300 from Patent WO0200928.
ACCESSION AX347229
VERSION AX347229.1 GI:18495117
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (sites)
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 2300 03-JAN-2002;
Epigenomics AG (DE)
FEATURES
source
1..1651
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 293 a 210 c 572 g 576 t
ORIGIN

Query Match 3.9%; Score 20; DB 6; Length 1651;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 gccgtgagggcgcgcggggg 161
|||||
Db 908 GCGCTGGAGGCGCGGGGGG 927
```

Search completed: September 20, 2002, 09:50:05
Job time: 12863 sec

Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgeson, A., Hoques, M., Holloway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E.,
 Jacobson, B., Jia, X., Johnson, R., Jolivet, S., Joudan, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lough, H.,
 Lozada, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokoko, S., Oguh, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoshitkar, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 175870)
 Worley, K. C.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

Submitted (22-MAY-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jan 31, 2002 this sequence version replaced gi:11415074.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HBBY
 Center clone name: RP11-461J7
 ----- Summary Statistics
 Sequencing vector: M13; L08821
 Chemistry: Dye-primer Bodypy; 21% of reads
 Chemistry: Dye-terminator Big Dye; 79% of reads
 Assembly program: Phrap; version 0.990329First call to
 findPhrapList

Consensus quality: 162338 bases at least Q40
 Consensus quality: 173354 bases at least Q30
 Consensus quality: 179527 bases at least Q20
 Estimated insert size: 176280; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-tp estimation
 Quality coverage: 3.4x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 19 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 36422: contig of 36422 bp in length
 * 36423 36522: gap of unknown length
 * 36523 63633: contig of 27111 bp in length
 * 63634 63733: gap of unknown length
 * 63734 74681: contig of 10948 bp in length
 * 74682 74781: gap of unknown length

74782 84412: contig of 9631 bp in length
 * 84413 84512: gap of unknown length
 * 84513 94505: contig of 9993 bp in length
 * 94506 94605: gap of unknown length
 * 94606 103746: contig of 9141 bp in length
 * 103747 103846: gap of unknown length
 * 112449 112449: contig of 8603 bp in length
 * 112450 112549: gap of unknown length
 * 112550 119806: contig of 7257 bp in length
 * 119807 119906: gap of unknown length
 * 119907 127705: contig of 7799 bp in length
 * 127706 127805: gap of unknown length
 * 127806 136906: contig of 9101 bp in length
 * 136907 137006: gap of unknown length
 * 137007 142959: contig of 5953 bp in length
 * 142960 143059: gap of unknown length
 * 143060 148560: contig of 5501 bp in length
 * 148561 148660: gap of unknown length
 * 148661 153887: contig of 5227 bp in length
 * 153888 153987: gap of unknown length
 * 153988 158883: contig of 4896 bp in length
 * 158884 158983: gap of unknown length
 * 158984 162866: contig of 3883 bp in length
 * 162867 162966: gap of unknown length
 * 162967 167882: contig of 4916 bp in length
 * 167883 167982: gap of unknown length
 * 167983 170496: contig of 2514 bp in length
 * 170497 170596: gap of unknown length
 * 170597 172912: contig of 2316 bp in length
 * 172913 173012: gap of unknown length
 * 173013 175870: contig of 2858 bp in length.

FEATURES
 Location/Qualifiers
 1..175870
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-461J7"

BASE COUNT 52311 a 35154 c 35502 g 51079 t 1824 others
 ORIGIN

Query Match 4.1% Score 21; DB 2; Length 175870;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 tgaactgtttacctaagg 376
 ||||||||||||||||
 Db 105138 TGAACCTGTGTACCTAATGG 105158

RESULT 14
 HSBHHV3
 LOCUS
 DEFINITION Bovine herpesvirus-1 DNA, fusion of left and right genomic termini.
 ACCESSION M19743
 VERSION M19743.1 GI:330677
 KEYWORDS
 SOURCE
 ORGANISM
 ORGANISM
 Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 Alphaherpesvirinae; Varicellovirus.
 1 (bases 1 to 210)
 Hammerschmidt, W., Ludwig, H. and Buhk, H.-J.
 Specificity of cleavage in replicative-form DNA of bovine
 herpesvirus 1
 J. Virol. 62, 1355-1363 (1988)
 MEDLINE
 88155775
 Location/Qualifiers
 1..210
 /organism="Bovine herpesvirus 1"
 /db_xref="taxon:10320"
 repeat_region 98..115

```

Qy 356 tgaactgtttaccctaatgg 376
|||||
Db 45731 TGAACCTGTTTACCTAATGG 45711

RESULT 11
AC005044
LOCUS AC005044 76714 bp DNA linear HTG 10-SEP-2000
DEFINITION Homo sapiens chromosome 7 clone CTB-4N9, WORKING DRAFT SEQUENCE, 5
unordered pieces.
ACCESSION AC005044
VERSION AC005044.3 GI:10048126
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 76714)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 76714)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 10, 2000 this sequence version replaced gi:9838351.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_RG004N09
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-primer ET; 77% of reads
Chemistry: Dye-terminator Big Dye; 23% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 67493 bases at least Q40
Consensus quality: 72535 bases at least Q30
Consensus quality: 75395 bases at least Q20
Insert size: 537000; agarose-fp
Insert size: 89515; sum-of-contigs
Quality coverage: 3.24 in Q20 bases; agarose-fp
Quality coverage: 3.66 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1206: contig of 1206 bp in length
* 1207 1306: gap of unknown length
* 1307 5096: contig of 3790 bp in length
* 5097 5196: gap of unknown length
* 5197 10233: contig of 5037 bp in length
* 10234 10333: gap of unknown length
* 10334 21887: contig of 11553 bp in length
* 21887 21987: gap of unknown length
* 21987 76714: contig of 54728 bp in length.
Location/Qualifiers
1 .76714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="CTB-4N9"

```

```

misc_feature 1..1206
/Note="assembly_name:Contig19
clone_end:SP6
vector_side:right"
misc_feature 1307..5096
/Note="assembly_name:Contig20"
misc_feature 5197..10233
/Note="assembly_name:Contig21"
misc_feature 10334..21886
/Note="assembly_name:Contig22"
misc_feature 21987..76714
/Note="assembly_name:Contig23"
BASE COUNT 22761 a 16021 c 14650 g 22878 t 404 others
ORIGIN

Query Match 4.1%; Score 21; DB 2; Length 76714;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 356 tgaactgtttaccctaatgg 376
|||||
Db 46933 TGAACCTGTTTACCTAATGG 46953

RESULT 12
AC005079_3
WPCOMMENT
Sequence split into 5 fragments LOCUS AC005079 Accession AC005079
Fragment Name Begin End
AC005079_0 1 110000
AC005079_1 100001 210000
AC005079_2 200001 310000
AC005079_3 300001 410000
AC005079_4 400001 415268
Continuation (4 of 5) of AC005079 from base 300001 (AC005079 Homo sapiens BAC clone C

Query Match 4.1%; Score 21; DB 9; Length 110000;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 356 tgaactgtttaccctaatgg 376
|||||
Db 50386 TGAACCTGTTTACCTAATGG 50406

RESULT 13
AC069221
LOCUS AC069221 175870 bp DNA linear HTG 31-JAN-2002
DEFINITION Homo sapiens chromosome 3 clone RP11-461J7, WORKING DRAFT SEQUENCE,
19 unordered pieces.
ACCESSION AC069221
VERSION AC069221.13 GI:18449557
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 175870)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,

```

Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 7696)
 Worley,K.C.
 Direct Submission
 Submitted (27-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GPMO
 Center clone name: CH230-151B7
 ----- Summary Statistics
 Sequencing vector: Plasmid; M77789
 Chemistry: Dye-terminator Big Dye; 92% of reads
 Assembly program: Phrap; version 0.990329First call to findPhrapList
 Consensus quality: 3425 bases at least Q40
 Consensus quality: 3898 bases at least Q30
 Consensus quality: 4368 bases at least Q20
 Estimated insert size: 1226; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 0x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 883: contig of 883 bp in length
 * 884 883: gap of unknown length
 * 984 1859: contig of 876 bp in length
 * 1860 1959: gap of unknown length
 * 1960 3095: contig of 1136 bp in length
 * 3096 3195: gap of unknown length
 * 3196 3788: contig of 593 bp in length
 * 3789 3888: gap of unknown length
 * 3889 5294: contig of 1406 bp in length
 * 5295 5394: gap of unknown length
 * 5395 6751: contig of 1357 bp in length
 * 6752 6851: gap of unknown length
 * 6852 7696: contig of 845 bp in length.

FEATURES
 source
 1..7696
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-151B7"
 BASE COUNT 1908 a 1929 c 1303 g 1932 t 624 others
 ORIGIN

Query Match 4.8%; Score 25; DB 2; Length 7696;
 Best Local Similarity 100.0%; Pred. No. 0.0029;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 atctctatgatgggaagccctaaga 348
 |||||
 Db 3584 ATCTCTATGATGGGAGCCCTAAGA 3608

RESULT 10
 AC005085/c
 LOCUS
 DEFINITION
 AC005085
 VERSION
 AC005085.2 GI:4415964
 KEYWORDS
 HTG; HTGS_PHASE1.
 SOURCE
 ORGANISM
 Homo sapiens
 Homo sapiens clone RG302D02, *** SEQUENCING IN PROGRESS ***, 15
 unordered pieces.
 AC005085
 VERSION
 AC005085.2 GI:4415964
 KEYWORDS
 HTG; HTGS_PHASE1.
 SOURCE
 ORGANISM
 Homo sapiens
 Homo sapiens clone RG302D02, *** SEQUENCING IN PROGRESS ***, 15
 unordered pieces.
 AC005085
 VERSION
 AC005085.2 GI:4415964
 KEYWORDS
 HTG; HTGS_PHASE1.
 SOURCE
 ORGANISM
 Homo sapiens
 Homo sapiens clone RG302D02, *** SEQUENCING IN PROGRESS ***, 15
 unordered pieces.

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

COMMENT

On Mar 13, 1999 this sequence version replaced gi:3212906.
 * NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2573: contig of 2573 bp in length
 * 2574 2585: gap of unknown length
 * 2586 5556: contig of 2971 bp in length
 * 5557 5568: gap of unknown length
 * 5569 7799: contig of 2231 bp in length
 * 7800 7811: gap of unknown length
 * 7812 10003: contig of 2192 bp in length
 * 10004 10015: gap of unknown length
 * 10016 12540: contig of 2525 bp in length
 * 12541 12552: gap of unknown length
 * 12553 18192: contig of 5640 bp in length
 * 18193 18204: gap of unknown length
 * 18205 22043: contig of 3839 bp in length
 * 22044 22055: gap of unknown length
 * 22056 29100: contig of 7045 bp in length
 * 29101 29112: gap of unknown length
 * 29113 33075: contig of 3963 bp in length
 * 33076 33087: gap of unknown length
 * 33088 36468: contig of 3381 bp in length
 * 36469 36480: gap of unknown length
 * 36481 38853: contig of 2373 bp in length
 * 38854 38865: gap of unknown length
 * 38866 44407: contig of 5542 bp in length
 * 44408 44419: gap of unknown length
 * 44420 47400: contig of 2981 bp in length
 * 47401 47412: gap of unknown length
 * 47413 53124: contig of 5712 bp in length
 * 53125 53135: gap of unknown length
 * 53136 58780: contig of 5645 bp in length.

FEATURES
 source
 1..58780
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RG302D02"
 BASE COUNT 17900 a 11129 c 11200 g 18272 t 279 others
 ORIGIN

Query Match 4.1%; Score 21; DB 2; Length 58780;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

* 26984 27912: contig of 929 bp in length
* 27913 28012: gap of 100 bp
* 28919: contig of 907 bp in length
* 28920 29019: gap of 100 bp
* 29020 29933: contig of 914 bp in length
* 29934 30033: gap of 100 bp
* 30930: contig of 897 bp in length
* 30931 31030: gap of 100 bp
* 31031 31924: contig of 894 bp in length
* 31925 32024: gap of 100 bp
* 32025 32902: contig of 878 bp in length
* 32903 33002: gap of 100 bp
* 33003 33890: contig of 888 bp in length
* 33891 33990: gap of 100 bp
* 33991 34900: contig of 910 bp in length
* 34901 35000: gap of 100 bp
* 35001 35895: contig of 895 bp in length
* 35896 35995: gap of 100 bp
* 35996 36840: contig of 845 bp in length
* 36841 36940: gap of 100 bp
* 36941 37822: contig of 882 bp in length
* 37823 37922: gap of 100 bp
* 37923 38845: contig of 923 bp in length
* 38846 38945: gap of 100 bp
* 38946 39823: contig of 878 bp in length
* 39824 39923: gap of 100 bp
* 39924 40792: contig of 869 bp in length
* 40793 40892: gap of 100 bp
* 40893 41811: contig of 919 bp in length
* 41812 41911: gap of 100 bp
* 41912 42806: contig of 895 bp in length
* 42807 42906: gap of 100 bp
* 42907 43784: contig of 878 bp in length
* 43785 43884: gap of 100 bp
* 43885 44777: contig of 893 bp in length
* 44778 44877: gap of 100 bp
* 44878 45768: contig of 891 bp in length
* 45769 45868: gap of 100 bp
* 45869 46761: contig of 893 bp in length
* 46762 46861: gap of 100 bp
* 46862 47713: contig of 852 bp in length
* 47714 47813: gap of 100 bp
* 47814 48729: contig of 916 bp in length
* 48730 48829: gap of 100 bp
* 48830 49779: contig of 950 bp in length
* 49780 49879: gap of 100 bp
* 49880 50794: contig of 915 bp in length
* 50795 50894: gap of 100 bp
* 50895 51797: contig of 903 bp in length
* 51798 51897: gap of 100 bp
* 51898 52786: contig of 889 bp in length
* 52787 52886: gap of 100 bp
* 52887 53779: contig of 893 bp in length
* 53780 53879: gap of 100 bp
* 53880 54782: contig of 903 bp in length
* 54783 54882: gap of 100 bp
* 54883 55769: contig of 887 bp in length
* 55770 55869: gap of 100 bp
* 55870 56736: contig of 867 bp in length
* 56737 56836: gap of 100 bp
* 56837 57715: contig of 879 bp in length
* 57716 57815: gap of 100 bp
* 57816 58712: contig of 897 bp in length
* 58713 58812: gap of 100 bp
* 58813 59723: contig of 917 bp in length
* 59730 59829: gap of 100 bp
* 59830 60638: contig of 809 bp in length
* 60639 60738: gap of 100 bp
* 60739 61654: contig of 916 bp in length
* 61655 61754: gap of 100 bp
* 61755 62680: contig of 926 bp in length
* 62681 62780: gap of 100 bp
* 62781 63664: contig of 884 bp in length

* 63665 63764: gap of 100 bp
* 63765 64670: contig of 906 bp in length
* 64671 64770: gap of 100 bp
* 64771 65666: contig of 896 bp in length
* 65667 65766: gap of 100 bp
* 65767 66636: contig of 870 bp in length
* 66637 66736: gap of 100 bp
* 66737 67604: contig of 868 bp in length
* 67605 67704: gap of 100 bp
* 67705 68576: contig of 872 bp in length
* 68577 68676: gap of 100 bp
* 68677 69601: contig of 925 bp in length
* 69602 69701: gap of 100 bp
* 69702 70618: contig of 917 bp in length

Query Match 5.0%; Score 26; DB 2; Length 87401;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 122 tqcaaggctgcgcctgcgcgcgtg 147
|||||
Db 5459 TCAAGGCTGCCCTGCCCGCGTG 5434
|||||

RESULT 9
AC108334
LOCUS
DEFINITION Rattus norvegicus clone CH230-151B7, *** SEQUENCING IN PROGRESS
AC108334 7696 bp DNA linear HTG 27-JAN-2002
***, 7 unordered pieces.
ACCESSION AC108334.1 GI:18377114
VERSION HTG; HTGS_PHASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 7696)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbaria, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buahy, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flag, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsso, F., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louissegh, H.,
Lozato, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okunolu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,

```

FEATURES
  source
    Location/Qualifiers
      1. .227144
        /organism="Mus musculus"
        /strain="C57BL6/J"
        /db_xref="taxon:10090"
        /chromosome="5"
        /clone_lib="RPC1 mouse BAC library 23"
        /clone="RP23-261D18"
      1. .26708
        /note="assembly_fragment"
        clone_end:SP6
        vector_side:left"
      26809..109655
        /note="assembly_fragment"
        clone_end:T7
        vector_side:left"
      109756..227144
        /note="assembly_fragment"
      68114 a 46378 c 46791 g 65638 t 223 others
      ORIGIN

  misc_feature
    Query Match          5.6%; Score 29; DB 2; Length 227144;
    Best Local Similarity 100.0%; Pred. No. 1.7e-05;
    Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  misc_feature
    QY 324 atctctatgatggagccctaaagccct 352
    Db 128178 ATCTCTATGATGGAGCCCTAAGAGCCT 128150

  RESULT 8
  AC021898/c
  LOCUS
  DEFINITION
    Homo sapiens chromosome 15 clone RP11-407J8 map 15, LOW-PASS
  SEQUENCE SAMPLING.
  AC021898
  VERSION
    AC021898.2 GI:9119282
  KEYWORDS
    HTG; HTGS_PHASE0.
  SOURCE
    human.
  ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 87401)
  AUTHORS
    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  JOURNAL
    Homo sapiens chromosome 15, clone RP11-407J8
  REFERENCE
    2 (bases 1 to 87401)
  AUTHORS
    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
    Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
    Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
    Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
    DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
    Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
    Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
    Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
    Rowland,J.C., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
    Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
    McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
    Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
    Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
    Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
    Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
    Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
    Zimmer,A. and Zody,M.
  DIRECT SUBMISSION
  TITLE
    Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
    Research, 320 Charles Street, Cambridge, MA 02141, USA
  JOURNAL
    On Jul 13, 2000 this sequence version replaced gi:6730901.
  COMMENT
    All repeats were identified using RepeatMasker:
    Smit, A.F.A. & Green, P. (1996-1997)
    http://ftp.genome.washington.edu/RN/RepeatMasker.html
    ----- Genome Center

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 12445
Center clone name: 407_J_8

* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
902 1001: gap of 100 bp
1002 1907: contig of 906 bp in length
1908 2007: gap of 100 bp
2008 2912: contig of 905 bp in length
2913 3012: gap of 100 bp
3013 3872: contig of 860 bp in length
3873 3972: gap of 100 bp
3973 4868: contig of 896 bp in length
4869 4968: gap of 100 bp
4969 5874: contig of 906 bp in length
5875 5974: gap of 100 bp
5975 6897: contig of 923 bp in length
6898 6997: gap of 100 bp
6998 7903: contig of 906 bp in length
7904 8003: gap of 100 bp
8004 8913: contig of 910 bp in length
8914 9013: gap of 100 bp
9014 9906: contig of 893 bp in length
9907 10006: gap of 100 bp
10007 10906: contig of 900 bp in length
10907 11006: gap of 100 bp
11007 11910: contig of 904 bp in length
11911 12010: gap of 100 bp
12011 12924: contig of 914 bp in length
12925 13024: gap of 100 bp
13025 13930: contig of 906 bp in length
13931 14030: gap of 100 bp
14031 14919: contig of 889 bp in length
14920 15019: gap of 100 bp
15020 15900: contig of 881 bp in length
15901 16000: gap of 100 bp
16001 16938: contig of 938 bp in length
16939 17038: gap of 100 bp
17039 17953: contig of 915 bp in length
17954 18053: gap of 100 bp
18054 18908: contig of 855 bp in length
18909 19008: gap of 100 bp
19009 19906: contig of 898 bp in length
19907 20006: gap of 100 bp
20007 20922: contig of 916 bp in length
20923 21022: gap of 100 bp
21023 21936: contig of 914 bp in length
21937 22036: gap of 100 bp
22037 22942: contig of 906 bp in length
22943 23042: gap of 100 bp
23043 23928: contig of 886 bp in length
23929 24028: gap of 100 bp
24029 24911: contig of 883 bp in length
24912 25011: gap of 100 bp
25012 25887: contig of 876 bp in length
25888 25987: gap of 100 bp
25988 26883: contig of 896 bp in length
26884 26983: gap of 100 bp

Assembly program: Phrap; version 0.990319
 Consensus quality: 199516 bases at least Q40
 Consensus quality: 200596 bases at least Q30
 Consensus quality: 201000 bases at least Q20
 Insert size: 197000; agarose-fp
 Insert size: 192000; pulse-field-gel
 Insert size: 204185; sum-of-contigs
 Quality coverage: 9.36x in Q20 bases; agarose-fp
 Quality coverage: 9.60x in Q20 bases; pulse-field-gel
 Quality coverage: 9.03x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 4428: contig of 4428 bp in length
 * 4429
 * 4528: gap of unknown length
 * 4529
 * 9553: contig of 5025 bp in length
 * 9554
 * 22746: contig of 13093 bp in length
 * 22747
 * 22846: gap of unknown length
 * 22847
 * 33954: contig of 11108 bp in length
 * 33955
 * 34054: gap of unknown length
 * 34055
 * 47831: contig of 13777 bp in length
 * 47832
 * 47931: gap of unknown length
 * 47932
 * 62010: contig of 14079 bp in length
 * 62011
 * 81789: contig of 19679 bp in length
 * 81790
 * 81889: gap of unknown length
 * 81890
 * 109919: contig of 28030 bp in length
 * 109920
 * 110019: gap of unknown length
 * 110020
 * 153235: contig of 43216 bp in length
 * 153236
 * 153335: gap of unknown length
 * 153336 205085: contig of 51750 bp in length.

FEATURES

FEATURES	Source
1..205085	/organism="Mus musculus"
	/strain="C57BL6/J"
	/db_xref="taxon:10090"
	/chromosome="5"
	/clone.lib="RPC1 mouse BAC library 23"
1..4428	
/note="assembly_fragment"	
4529..9553	
/note="assembly_fragment"	
9654..22746	
/note="assembly_fragment"	
22847..33954	
/note="assembly_fragment"	
34055..47831	
/note="assembly_fragment"	
clone_end:T7	
vector_side:right"	
47932..62010	
/note="assembly_fragment"	
62111..81789	
/note="assembly_fragment"	
81890..109919	
/note="assembly_fragment"	
110020..153235	
/note="assembly_fragment"	
153336..205085	
/note="assembly_fragment"	
clone_end:SP6	
vector_side:left"	
61131 a 40716 c 41296 g 61026 t 916 others	
BASE COUNT	
ORIGIN	

Query Match 5.68; Score 29; DB 2; Length 205085;
 Best Local Similarity 100.0%; Pred. No. 1.6e-05;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 324 atctctatgatggagccctaaagcct 352
 |||||
 Db 186665 ATCTCTATGATGGAGCCCTAAGGCCT 186693

RESULT

7
 AC092404/c
 LOCUS 227144 bp DNA linear HTG 04-JUL-2001
 DEFINITION Mus musculus chromosome 5 clone RP23-261D18 strain C57BL6/J,
 WORKING DRAFT SEQUENCE, 3 unordered pieces.
 AC092404
 ACCESSION
 VERSION AC092404.1 GI:14595778
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 227144)
 Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
 Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L.,
 Granite,S., Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E.,
 Lee-Lin,S.-O., Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B.,
 Masiello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
 Pearson,R., Prasad,A., Shevchenko,Y., Snyder,B., Stantirpop,S.,
 Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L.,
 Walker,M.A., Wetherby,K.D., Zhang,L.-H. and Green,E.D.
 NISC Comparative Sequencing Initiative
 Unpublished
 2 (bases 1 to 227144)
 Green,E.D.
 Direct Submission
 Submitted (04-JUL-2001) NIH Intramural Sequencing Center, 8717
 Grovemont Circle, Gaithersburg, MD 20877, USA
 ----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: http://www.nisc.nih.gov
 Contact: nisc.mouse@nih.gov
 ----- Project Information
 Center project name: qv
 Center clone name: 261D18
 ----- Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 225678 bases at least Q40
 Consensus quality: 225959 bases at least Q30
 Consensus quality: 226087 bases at least Q20
 Insert size: 206000; agarose-fp
 Insert size: 225000; pulse-field-gel
 Insert size: 226944; sum-of-contigs
 Quality coverage: 14.02x in Q20 bases; agarose-fp
 Quality coverage: 12.84x in Q20 bases; pulse-field-gel
 Quality coverage: 12.73x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 26708: contig of 26708 bp in length
 * 26709
 * 26808: gap of unknown length
 * 26809 109655: contig of 82847 bp in length
 * 109656 109755: gap of unknown length
 * 109756 227144: contig of 117389 bp in length.

```

QY 167 agaaatgaagagagaaattggaataattgtagtgagacattctgtactgtctctcctt 226
Db 61930 AGAAGTGAAGAGAGAAATGGAAATTTGTGAGTGGACCTTCTGTACTACTGCTCCTCTT 61989

QY 227 gqgtgaaaggggaagactgcagtcattattcagc 266
Db 61990 GCGTGAAGAGGGAAGAACTGCATCATATATTTCAGC 62029

RESULT 5
AC099742
LOCUS AC099742 172915 bp DNA linear HTG 20-NOV-2001
DEFINITION Papio cynocephalus anubis clone RP41-167P22, WORKING DRAFT
AC099742
ACCESSION AC099742.1 GI:17017546
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS olive baboon.
SOURCE Papio cynocephalus anubis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopitheciinae; Papio.
REFERENCE 1 (bases 1 to 172915)
AUTHORS Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Shevchenko,Y.,
Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
Tsurgueon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 172915)
Green,E.D.
Direct Submission
Submitted (20-NOV-2001) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@hgri.nih.gov
----- Project Information
Center project name: ccy
Center clone name: 167P22
----- Summary Statistics
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171178 bases at least Q40
Consensus quality: 171910 bases at least Q30
Consensus quality: 172236 bases at least Q20
Insert size: 130000; agarose-fp
Insert size: 172615; sum-of-ctnigs
Quality coverage: 12.58x in Q20 bases; agarose-fp
Quality coverage: 9.47x in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2438: contig of 2438 bp in length
* 2539 8133: gap of unknown length
* 8134 8233: gap of 5595 bp in length
* 8234 40378: gap of unknown length
* 40379 40478: contig of 32145 bp in length
* 40478: gap of unknown length

```

```

FEATURES
source
* 40479 172915: contig of 132437 bp in length.
Location/Qualifiers
1..172915
/organism="Papio cynocephalus anubis"
/db_xref="taxon:9555"
/clone="RP41-167P22"
/clone_lib="RP41"
1..2438
/note="assembly_fragment"
clone_end:SP6
vector_side:right
misc_feature 2539..8133
/note="assembly_fragment"
8234..40378
/note="assembly_fragment"
clone_end:T7
vector_side:right
misc_feature 40479..172915
/note="assembly_fragment"
BASE COUNT 52509 a 31733 c 32277 g 56096 t 300 others
ORIGIN
Query Match 10.4%; Score 54; DB 2; Length 172915;
Best Local Similarity 100.0%; Pred. No. 7.6e-21;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 310 ctatcatgaatcaatctctatgtatggagccctaaagacgtttagtgaactt 363
|||||
Db 83218 GTATCATGGAATCAATCTCTATGATGGGAGCCCTTAGAGCCTTAGTGAACCTT 83271
|||||

RESULT 6
AC026813
LOCUS AC026813 205085 bp DNA linear HTG 27-OCT-2001
DEFINITION Mus musculus chromosome 5 clone RP23-119M19 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 10 unordered pieces.
AC026813
ACCESSION AC026813.2 GI:16506398
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 205085)
AUTHORS Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Shevchenko,Y.,
Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
Tsurgueon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 205085)
Green,E.D.
Direct Submission
Submitted (24-MAR-2000) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
On Oct 27, 2001 this sequence version replaced gi:7321459.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@hgri.nih.gov
----- Project Information
Center project name: xh
Center clone name: 119M19
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

```

* 1 901: contig of 901 bp in length
* 902 1001: gap of 100 bp
* 1002 1907: contig of 906 bp in length
* 1908 2007: gap of 100 bp
* 2008 2912: contig of 905 bp in length
* 2913 3012: gap of 100 bp
* 3013 3872: contig of 860 bp in length
* 3873 3972: gap of 100 bp
* 3973 4868: contig of 896 bp in length
* 4869 4968: gap of 100 bp
* 4969 5874: contig of 906 bp in length
* 5875 5974: gap of 100 bp
* 5975 6897: contig of 923 bp in length
* 6898 6997: gap of 100 bp
* 6998 7903: contig of 906 bp in length
* 7904 8003: gap of 100 bp
* 8004 8913: contig of 910 bp in length
* 8914 9013: gap of 100 bp
* 9014 9906: contig of 893 bp in length
* 9907 10006: gap of 100 bp
* 10007 10906: contig of 900 bp in length
* 11007 11006: gap of 100 bp
* 11007 11910: contig of 904 bp in length
* 11911 12010: gap of 100 bp
* 12011 12924: contig of 914 bp in length
* 12925 13024: gap of 100 bp
* 13025 13930: contig of 906 bp in length
* 13931 14030: gap of 100 bp
* 14031 14919: contig of 889 bp in length
* 14920 15019: gap of 100 bp
* 15020 15900: contig of 881 bp in length
* 15901 16000: gap of 100 bp
* 16001 16938: contig of 938 bp in length
* 16939 17038: gap of 100 bp
* 17039 17953: contig of 915 bp in length
* 17954 18053: gap of 100 bp
* 18054 18908: contig of 855 bp in length
* 18909 19008: gap of 100 bp
* 19009 19906: contig of 898 bp in length
* 19907 20006: gap of 100 bp
* 20007 20922: contig of 916 bp in length
* 20923 21022: gap of 100 bp
* 21023 21936: contig of 914 bp in length
* 21937 22036: gap of 100 bp
* 22037 22942: contig of 906 bp in length
* 22943 23042: gap of 100 bp
* 23043 23928: contig of 886 bp in length
* 23929 24028: gap of 100 bp
* 24029 24911: contig of 883 bp in length
* 24912 25011: gap of 100 bp
* 25012 25887: contig of 876 bp in length
* 25888 25987: gap of 100 bp
* 25988 26883: contig of 896 bp in length
* 26884 26983: gap of 100 bp
* 26984 27912: contig of 929 bp in length
* 27913 28012: gap of 100 bp
* 28013 28919: contig of 907 bp in length
* 28920 29019: gap of 100 bp
* 29020 29933: contig of 914 bp in length
* 29934 30033: gap of 100 bp
* 30034 30930: contig of 897 bp in length
* 30931 31030: gap of 100 bp
* 31031 31924: contig of 894 bp in length
* 31925 32024: gap of 100 bp
* 32025 32902: contig of 878 bp in length
* 32903 33002: gap of 100 bp
* 33003 33890: contig of 888 bp in length
* 33891 33990: gap of 100 bp
* 33991 34900: contig of 910 bp in length
* 34901 35000: gap of 100 bp
* 35001 35895: contig of 895 bp in length
* 35896 35995: gap of 100 bp
* 35996 36840: contig of 845 bp in length

* 36841 36940: gap of 100 bp
* 36941 37822: contig of 882 bp in length
* 37823 37922: gap of 100 bp
* 37923 38845: contig of 923 bp in length
* 38846 38945: gap of 100 bp
* 38946 39823: contig of 878 bp in length
* 39824 39923: gap of 100 bp
* 39924 40792: contig of 869 bp in length
* 40793 40892: gap of 100 bp
* 40893 41811: contig of 919 bp in length
* 41812 41911: gap of 100 bp
* 41912 42806: contig of 895 bp in length
* 42807 42906: gap of 100 bp
* 42907 43784: contig of 878 bp in length
* 43785 43884: gap of 100 bp
* 43885 44777: contig of 893 bp in length
* 44778 44877: gap of 100 bp
* 44878 45768: contig of 891 bp in length
* 45769 45868: gap of 100 bp
* 45869 46761: contig of 893 bp in length
* 46762 46861: gap of 100 bp
* 46862 47713: contig of 852 bp in length
* 47714 47813: gap of 100 bp
* 47814 48729: contig of 916 bp in length
* 48730 48829: gap of 100 bp
* 48830 49779: contig of 950 bp in length
* 49780 49879: gap of 100 bp
* 49880 50794: contig of 915 bp in length
* 50795 50894: gap of 100 bp
* 50895 51797: contig of 903 bp in length
* 51798 51897: gap of 100 bp
* 51898 52786: contig of 889 bp in length
* 52787 52886: gap of 100 bp
* 52887 53779: contig of 893 bp in length
* 53780 53879: gap of 100 bp
* 53880 54782: contig of 903 bp in length
* 54783 54882: gap of 100 bp
* 54883 55769: contig of 887 bp in length
* 55770 55869: gap of 100 bp
* 55870 56736: contig of 867 bp in length
* 56737 56836: gap of 100 bp
* 56837 57715: contig of 879 bp in length
* 57716 57815: gap of 100 bp
* 57816 58712: contig of 897 bp in length
* 58713 58812: gap of 100 bp
* 58813 59729: contig of 917 bp in length
* 59730 59829: gap of 100 bp
* 59830 60638: contig of 809 bp in length
* 60639 60738: gap of 100 bp
* 60739 61654: contig of 916 bp in length
* 61655 61754: gap of 100 bp
* 61755 62680: contig of 926 bp in length
* 62681 62780: gap of 100 bp
* 62781 63664: contig of 884 bp in length
* 63665 63764: gap of 100 bp
* 63765 64670: contig of 906 bp in length
* 64671 64770: gap of 100 bp
* 64771 65666: contig of 896 bp in length
* 65667 65766: gap of 100 bp
* 65767 66636: contig of 870 bp in length
* 66637 66736: gap of 100 bp
* 66737 67604: contig of 868 bp in length
* 67605 67704: gap of 100 bp
* 67705 68576: contig of 872 bp in length
* 68577 68676: gap of 100 bp
* 68677 69601: contig of 925 bp in length
* 69602 69701: gap of 100 bp
* 69702 70618: contig of 917 bp in length

Query Match 19.3%; Score 100; DB 2; Length 87401;
Best Local Similarity 100.0%; Pred. No. 4.4e-49;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: <http://www.nisc.nih.gov>
 Contact: nisc_mouse@nhgri.nih.gov
 ----- Project Information
 Center project name: ces
 Center clone name: 120K11
 ----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 161001 bases at least Q40
 Consensus quality: 161422 bases at least Q30
 Consensus quality: 161562 bases at least Q20
 Insert size: 143000; agarose-fp
 Insert size: 162428; sum-of-contigs
 Quality coverage: 10.76x in Q20 bases; agarose-fp
 Quality coverage: 9.47x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

* 1 13922: contig of 13922 bp in length
 * 13923 14022: gap of unknown length
 * 14023 36248: contig of 22226 bp in length
 * 36249 36348: gap of unknown length
 * 36349 66192: contig of 29844 bp in length
 * 66193 66292: gap of unknown length
 * 66293 92268: contig of 25876 bp in length
 * 92269 122036: gap of unknown length
 * 122037 122136: contig of 29768 bp in length
 * 122137 162928: contig of 40792 bp in length.

FEATURES

source
 1. 162928
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="RP43-120K11"
 /clone_lib="RP43"

misc_feature

1. 13922
 /note="assembly_fragment"
 clone_end:77
 vector_side:left

misc_feature

14023..36248
 /note="assembly_fragment"

misc_feature

36349..66192
 /note="assembly_fragment"

misc_feature

66293..92268
 /note="assembly_fragment"

misc_feature

92269..122036
 /note="assembly_fragment"

misc_feature

122137..162928
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:right

BASE COUNT 51528 a 29878 c 30533 g 50485 t 504 others
 ORIGIN

Query Match 26.9%; Score 139; DB 2; Length 162928;
 Best Local Similarity 99.5%; Prid. No. 6e-73;
 Matches 189; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 280 aggatatttggatcgttggaagtgcggtatcatgaatgaatcctctatgatggaa 339
 |||||||

DB 37913 AGGATATTCCTGTCCTGGAAGTCCGTCATCATGAATCAATCTCTATGATGGAA 37854
 |||||||

QY 340 gccctaagagccttagtgaacctgtttaccctaagtcataaagtgtatcaaatgacaa 399
 |||||||

Db 37853 GGCCTAAGAGCCTTAGTGAACCTTTTACCTAATGCCATAAATGCTATCAAGATGCAA 37794
 |||||||
 QY 400 ggaagtcacttagtggtggaagtggagattttgccaaatccttgaccattcgac 459
 |||||||
 Db 37793 GGAAGTCACTTAGTGTGATGGAAGTGGAGATTTTGCCTAATCTTGACCATCGAC 37734
 |||||||
 QY 460 ttattagatg 469
 |||||||
 Db 37733 TTATTAGATG 37724
 |||||||

RESULT 4

AC021898
 Homo sapiens chromosome 15 clone RP11-407J8 map 15, linear HTG 13-JUL-2000
 DEFINITION
 SEQUENCE SAMPLING.

AC021898.2 GI:9119282

AC021898

HTG: HTGS_PHASE0.

human

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 87401)

Birren.B., Linton.L., Nusbaum.C. and Lander.E.

Homo sapiens chromosome 15, clone RP11-407J8

Unpublished

2 (bases 1 to 87401)

Birren.B., Linton.L., Nusbaum.C., Lander.E., Abraham.H., Allen.N.,

Anderson.S., Baldwin.J., Barna.N., Beckerly.R., Beda.F.,

Boguslavsky.L., Boukhgaiter.B., Brown.A., Burkett.G., Castle.A.,

Chepel.Y., Colangelo.M., Collins.S., Collamore.A., Cooke.P.,

DeArelano.K., Dewar.K., Domino.M., Doyle.M., Fenestor.J.,

Ferreira.P., FitzHugh.W., Forrest.C., Gage.D., Galagan.J.,

Gerdyne.S., Grant.G., Hagos.B., Heaford.A., Horton.L.,

Howland.J.C., Johnson.R., Jones.C., Kann.L., Karatas.A., Klein.J.,

Landers.T., Lechoczky.J., Levine.R., Lieu.C., Liu.G., Locke.K.,

Macdonald.P., Marquis.N., McEwan.P., McGurk.A., McKernan.K.,

McPheeters.R., Meldrim.J., Meneus.L., Morrow.J., Naylor.J.,

Norman.C.H., O'Connor.T., O'Donnell.P., Olivari.T.M., Peterson.K.,

Pierre.N., Pisani.C., Pollara.V., Raymond.C., Riley.R., Rothman.D.,

Roy.A., Santos.R., Severy.P., Spencer.B., Stange-Thomann.N.,

Stojanovic.N., Subramanian.A., Talamas.J., Testaye.S., Theodore.J.,

Tirrell.A., Vassiliev.H., Viel.R., Vo.A., Wu.X., Wyman.D., Ye.W.J.,

Zimmer.A. and Zody.M.

Direct Submission

Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6730901.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2445

Center clone name: 407_J_8

* NOTE: This record contains 88 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

/note="similar to human EST AA151807 (NID:g1720502)
z139g02.rl"
27125..27544
/note="similar to human EST AA151796 (NID:g1720491)
z139c02.rl"
complement(27178..27600)
/note="similar to human EST AA149579 (NID:g1720380)
z139c02.sl"
complement(27441..27890)
/note="similar to human EST N52554 (NID:g1193720)
yv36a11.sl"
27686..27856
/note="similar to human EST W32120 (NID:g1313113)
zb97c08.rl"
27686..27838
/note="similar to human EST W32085 (NID:g1313105)
zb97b08.rl"
complement(27733..28143)
/note="similar to human EST N59831 (NID:g1203721)
yz77a10.sl"
complement(27838..28143)
/note="similar to human EST W31561 (NID:g1312680)
zb97b08.sl"
complement(27887..28143)
/note="similar to human EST N59830 (NID:g1203720)
yz77a09.sl"
complement(27997..28144)
/note="similar to human EST W31628 (NID:g1312688)
zb97c08.sl"
28666..28972
/note="similar to human EST R09339 (NID:g761262)
yf26d05.rl"
complement(29587..29853)
/note="similar to human EST R09227 (NID:g761150)
yf26d05.sl"
30146..30314
/note="similar to human EST N78042 (NID:g1240743)
yv71e06.rl"
complement(31170..31350)
/note="similar to human EST N58451 (NID:g1202341)
yv71e06.sl"
31970..32006
/rpt_family="L1"
32263..32563
/rpt_family="ALU"
complement(34737..34778)
/rpt_family="L1"
36558..36709
/rpt_family="ALU"
complement(37630..37740)
/rpt_family="PTR"
39839..40125
/rpt_family="ALU"
40126..40157
/rpt_family="L1"
complement(41340..41706)
/rpt_family="ALU"
complement(41904..41933)
/rpt_family="L1"
42348..42409
/rpt_family="ALU"
complement(42942..42982)
/rpt_family="L1"
complement(44380..44423)
/rpt_family="L1"
44659..44954
/rpt_family="ALU"
45422..45448
/rpt_family="L1"
complement(45615..45669)
/rpt_family="L1"
complement(45672..45963)
/rpt_family="ALU"

repeat_region complement(45964..46597)
/rpt_family="L1"
repeat_region 46100..46516
/rpt_family="L1"
repeat_region 48370..48785
/rpt_family="L1"
repeat_region complement(49577..50216)
/rpt_family="L1"
repeat_region complement(50540..50791)
/rpt_family="THE"
repeat_region complement(50820..50985)
/rpt_family="THE"
repeat_region complement(51124..51156)
/rpt_family="L1"
repeat_region 51405..51691
/rpt_family="ALU"
repeat_region complement(51721..51764)
/rpt_family="L1"
repeat_region complement(51765..52058)
/rpt_family="ALU"
repeat_region complement(52252..52467)
Query Match 34.48; Score 178; DB 9; Length 156214;
Best Local Similarity 99.68; Pred.No.7.5e-97;
Matches 228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 280 aggatattcttggtgattggaagatgctcgatcatcgaatcaatctctatgaggaa 339
|||||
Db 17321 AGGATATTCCTGGTGATCTTGGAAAGTGTCGTATCATCGAATCAATCTCTATGAGGAA 17380
Qy 340 gccctaagagccttagtgaaactgtttaccataatggcataaaatggatcaaatgcaaa 399
|||||
Db 17381 GCCCTAAGAGCCTTAGTGAAACTTTTACCTAATGTCATAAATGGTATCAAGATGCAA 17440
Qy 400 ggaaggctcaactgtaggtgattggaagtgagattttgccaatccttgaccattgcac 459
|||||
Db 17441 GGAAGGTCACCTGTAGGTGATTTGGAAGTGGAGATTTTGCAGAAATCTTGACCATTCGAC 17500
Qy 460 ttattagatcggtcatcatgctgctcataggaagtagaatacctaaagt 508
|||||
Db 17501 TTATTAGATGGGCTATCATGTGCTCATAGGAAGTAGAATCTTAAGTT 17549
RESULT 3
AC104475/c
LOCUS
DEFINITION Pan troglodytes clone RP43-120K11, WORKING DRAFT SEQUENCE, 6
unordered pieces.
AC104475
AC104475.1 GI:17530717
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE chimpanzee.
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
1 (bases 1 to 162928)
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-Q.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Masello, C., Mastrian, S.D.,
McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Stantripoop, S.,
Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L.,
Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and
Green, E.D.
NISC Comparative Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 162928)
Green, E.D.
AUTHORS
Direct Submission
TITLE
Submitted (12-DEC-2001) NIH Intramural Sequencing Center, 8717
Gommet Circle, Gaithersburg, MD 20877, USA
JOURNAL
COMMENT
----- Genome Center

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

FROM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 06:15:42 ; Search time 4356.39 Seconds
(without alignments)
2483.483 Million cell updates/sec

Title: US-09-802-520-3
Perfect score: 517
Sequence: 1 qggaaagcagctggagtgcg.....aatcctaaattgagcttctga 517

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

```
Word size : 0
Total number of hits satisfying chosen parameters: 3595312
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

Database : GenEmbl:*

```

1: gb_ba:*
2: gb_h9:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
--------	-------	-------	-------	--------	----	----	-------------

ALIGNMENTS

RESULT	1
AXI55249	
LOCUS	linear PAT 22-JUN-2001
DEFINITION	Sequence 7 from Patent WO0140276.
ACCESSION	AXI55249
VERSION	AXI55249.1 GI:14536713
KEYWORDS	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 2453)
AUTHORS	Afar, D.E., Hubert, R.S., Raitano, A.B., Saffran, D.C., Mitchell, S.C., Faris, M. and Jakobovits, A.
TITLE	Serpentine transmembrane antigens expressed in human prostate cancers and uses thereof
JOURNAL	Patient: WO 0140276-A 7 07-JUN-2001;
FEATURES	Urogenesys, Inc. (US) Location/Qualifiers 1..2453 /organism="Homo sapiens" /db_xref="taxon:9606"
source	355...1719
CDS	